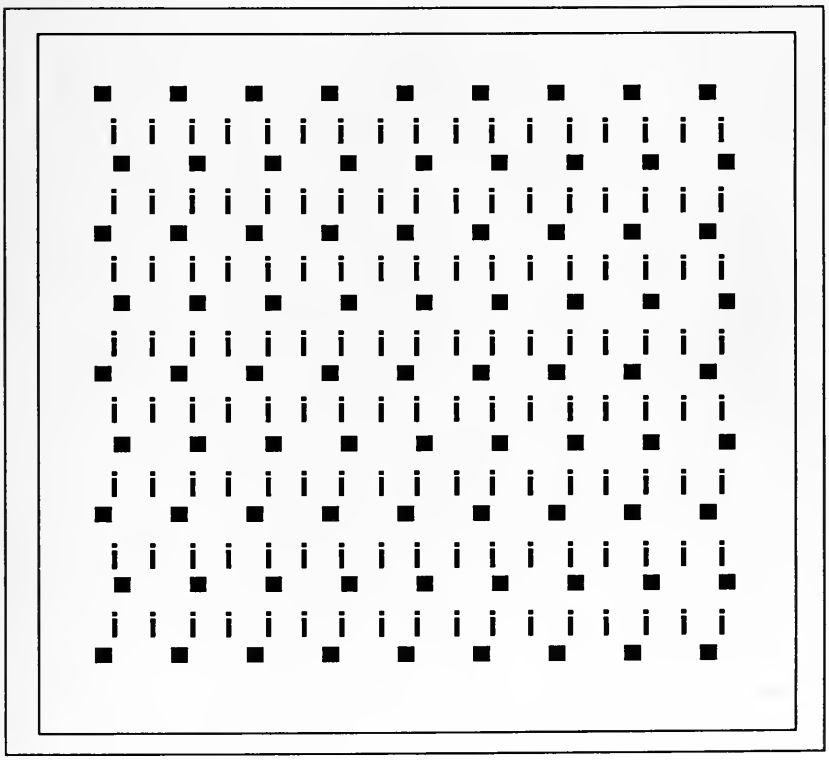
IASSIST

QUARTERLY

VOLUME 15 Summer 1991 NUMBER 2



Printed at UCLA

Digitized by the Internet Archive
in 2010 with funding from
University of North Carolina at Chapel Hill

http://www.archive.org/details/iassistquarterly152inte

IASSIST QUARTERLY

The IASSIST QUARTERLY represents an international cooperative effort on the part of individuals managing, operating, or using machine-readable data archives, data libraries, and data services. The QUARTERLY reports on activities related to the production, acquisition, preservation, processing, distribution, and use of machine-readable data carried out by its members and others in the international social science community. Your contributions and suggestions for topics of interest are welcomed. The views set forth by authors of articles contained in this publication are not necessarily those of IASSIST.

Information for Authors

The QUARTERLY is published four times per year. Articles and other information should be typewritten and double-spaced. Each page of the manuscript should be numbered. The first page should contain the article title, author's name, affiliation, address to which correspondence may be sent, and telephone number. Footnotes and bibliographic citations should be consistent in style, preferably following a standard authority such as the University of Chicago press *Manual of Style* or Kate L. Turabian's *Manual for Writers*. Where appropriate, machine-readable data files should be cited with bibliographic citations consistent in style with Dodd, Sue A. "Bibliographic references for numeric social science data files: suggested guidelines". *Journal of the American Society for Information Science 30(2):77-82, March 1979*. If the contribution is an announcement of a conference, training session, or the like, the text should include a mailing address and a telephone number for the director of the event or for the organization sponsoring the event. Book notices and reviews should not exceed two double-spaced pages. Deadlines for submitting articles are six weeks before publication. Manuscripts should be sent in duplicate to the Editor: Walter Piovesan, Research Data Library, W.A.C. Bennett Library, Simon Fraser University, Burnaby, B.C., V5A 1S6 CANADA. (604) 291-4349 E-Mail: USERDLIB@SFU.BITNET Book reviews should be submitted in duplicate to the Book Review Editor: Daniel Tsang, Main Library, University of California P.O. Box 19557, Irvine, California 92713 USA. (714) 856-4978 E-Mail: DTSANG@ORION.CF.UCI.EDU

Title: Newsletter - International Association for Social Science Information Service and Technology
ISSN - United States: 0739-1137 Copyright 1985 by IASSIST. All rights reserved.

CONTENTS

Volume 15 Number 2 Summer 1991

FEATURES

4 **Implementation of CATI Techniques in an Academic Social Science Research Setting**
by Dave Odynak and Cliff Kinzel

10 **Evaluation and Appraisal of University Administrative Computing Datasets**
by Mark Conrad

14 **Remote Access to Local Data Collections: A UC Davis Experiment**
by Jean Slemmons Stratford

16 **History and Statistical Analysis: A case study**
by Patricia E. Prestwich

20 **The U.S. Public Use Census Microdata Files as a Source for the Study of Long-Term Social Change**
by Steven Ruggles

29 **Integrating Data Resources and Library Resources: The SPIRES Experience**
by Slavko Manojlovich

News

38 **IASSIST '92 conference**

39 **IASSIST membership**

Implementation of CATI Techniques in an Academic Social Science Research Setting

by Dave Odynak [1]
and Cliff Kinzel
University of Alberta
Edmonton

Abstract
This paper focuses on the use of computer-assisted techniques for conducting survey interviews by telephone (CATI). A few of the basic and some of the more advanced features of CATI systems are briefly listed. Experiences with integrating the CATI electronic questionnaire into survey planning, training, execution and data processing in a academic research setting accustomed to traditional pencil and paper survey methods are discussed. Special attention is placed on the social aspects of introducing CATI technology into an established research environment.

Experience Meets Technology
Computer assisted telephone interviewing techniques (CATI) have become an extremely important consideration to researchers using the telephone interviewing mode of data collection. The Population Research Laboratory (PRL) has been actively engaged in survey research since 1973 and has administered numerous survey projects through its facilities. These survey projects have involved face-to-face, telephone, and mail-out questionnaires.

At the time of planning for a recent (1989) small survey project in the PRL using telephone interviewing there was very limited use of CATI in social science research centres in Canada. This perceived lack of knowledge and expertise meant that the integration of a CATI system into a research setting such as the PRL would ultimately rely on trial and error. Moreover, the CATI system would be considered as an extension of, rather than a replacement for the traditional mode of conducting surveys. After receiving funding, hiring a researcher for the survey, and then conducting a comparison of existing CATI systems for the microcomputer, the Population Research Laboratory decided to use a CATI system. The main use of the system would be for a small academic survey of public opinion in a localized area (province) rather than for a large national survey or commercial marketing application.

Surveys in an Academic Setting
The Population Research Laboratory functions as the research wing of the Department of Sociology at the University of Alberta. Because the PRL is located within an academic social science setting it must perform service roles within the University and the Department of Sociology and still run survey projects in a cost efficient manner. The introduction of new survey technologies into academic research settings such as the PRL is necessary to keep pace with the competition for research dollars from non-profit and commercial consulting organizations.

On the one hand, some flexibility is allowed within an academic setting for research into new survey methodologies. Clearly, surveys are conducted in an academic setting for purposes other than trying to make money. An examination of improvements in survey methodologies is encouraged and rewarded in the form of publications, research papers, and graduate theses. On the other hand, survey research can become overburdened with the numerous operating constraints that exist in large bureaucratic institutions like a university. For instance, survey projects are subject to university overhead costs, contract requirements, and public tender on goods. Moreover, conducting surveys in an academic setting can compete for goods and services required for other functions that a research centre like the PRL performs in a university. For example, in addition to providing research facilities for survey projects, the PRL also publishes an academic journal in population, provides staff reprints and discussion papers, consults, coordinates conferences and workshops for the Department and University, and provides staff and supervision of the Department reading room.

Pencil and Paper Techniques
For most surveys, the PRL staff rely heavily on pencil and paper techniques with the majority of the data processing done on the mainframe. Over time, technological improvements are slowly integrated into the survey process. Two examples of bringing new ideas into the survey process are: data entry done directly onto the mainframe replacing data keypunched on cards, and the use of optical scan sheets to collect household information at the start of the interview. However, the advent of CATI in the PRL meant substantial changes to the survey process in a relatively short time-span. This is an important consideration for survey research centres that have well-tested and established procedures for doing surveys.

The administration of the telephone questionnaire was closely modelled after the face-to-face procedures with some modifications. Both types of surveys are characterized by large stacks of paper questionnaires moving through the various stages of survey processing.

The stages in the PRL telephone survey process are described briefly in the following steps. Samples of respondents are drawn using probability sampling techniques (Kinzel 1989). Interviewers and supervisors are hired and trained. Questionnaires are pretested, modifications made and field work commenced. A modified probability sample emerges through random digit dialing techniques and quota sampling at the household level. The interview begins with a well-tested introduction that is relatively successful in gaining cooperation while maintaining informed consent. After an optical scan sheet with an accounting of household members is filled out, an interviewer flips through the paper questionnaire, follows the instructions, asks the questions and records manually on the paper questionnaire the respondent's answers. The questionnaires completed by the interviewers are then edited by supervisors or research assistants. At this stage some of the out-of-range and inconsistent responses are caught. Call-backs and validations checks by supervisors to the survey respondents are made to confirm and obtain additional information and these adjustments are also recorded on the paper form.

Following the completion of interviewing, which takes around two months, responses on the questionnaires for both close-ended and open-ended questions are coded and transferred onto IBM coding sheets. The codes on the sheets are then keypunched into a computer data file. An SPSSx command file is written to read this data file. At this stage, the data are cleaned again using special programs written in Fortran that flag inconsistencies and wildcodes. Typically, for a survey with 1,000 or more respondents and interviews approximately 40 minutes long, it might take up to three months for the coding, data entry, and cleaning to be completed before the data are ready for analysis.

CATI Capabilities

By contrast to the traditional pencil and paper methods previously described, CATI presents interviewers with an electronic questionnaire with the questions, instructions, and choices shown on a computer monitor and answers typed in via keyboard. Much of the editing for wildcodes and inconsistencies is built into the electronic questionnaire so that the cleaning of the data is done continually throughout the execution of the survey. Moreover, the recording of responses by interviewers represents automatic data entry.

Two types of CATI systems exist for microcomputers: CATI programs that operate on stand-alone computers and CATI programs that can take advantage of networked computers. CATI systems for either type are available commercially. Nicholls (1988) provides an excellent general introduction to the history, use and capabilities of CATI systems. Some notable features available on the commercial CATI stand-alone version purchased by the PRL are:

- o automatic generation of skip patterns or routing
- o handling of special question types including close-ended and open-ended multiple response
- o variable text insertion into questions based on previous responses
- o an electronic coder to code open-ended question responses on-line
- o randomization of questions and question/answer choice
- o arithmetic functions
- o on-line editing of data files
- o export of data into files for statistical analysis

Most of the above features duplicate the traditional pen and pencil questionnaire method using interactive computing to assist interviewers. Even more features are available in CATI systems that are networked. In the networked version of CATI some substantial changes added to the supervisory and administrative functions in the survey process are:

- o automatic monitoring of interviewer performance and survey indicators such as response rate, quotas and completions
- o call disposition monitoring and incidence reports
- o automatic scheduling of call-backs
- o integration of random digit dialing and call scheduling

Research methodologists at the PRL attempted to implement as many of the features of the new CATI

system as possible. The main reason for the thrust was an interest in future use of CATI technology for the telephone component of the PRL's annual omnibus-style survey. It was a major challenge for the PRL to adapt the CATI system to capitalize on the survey administration experience of the facilities' personnel and well-tested procedures for doing surveys. Another challenge was to use the available features of the CATI system to accommodate the many different question types and options that appear on a typical PRL survey.

Survey Planning

The planning of any survey can be an arduous task involving a multitude of details not readily apparent to a layperson. The survey project on public attitudes toward discharged psychiatric patients was no exception. Three separate sub-samples across the province of Alberta were drawn and contacted using random digit dialing techniques and then interviewed. Many of the questions were replicated from the 1986 Winnipeg Area Study which had employed face-to-face techniques (Currie 1986).

Most survey tasks were envisioned and listed before the purchase of the CATI software. After the software arrived several additional steps were added to the planning list but there was still a great deal to learn about the CATI system during the actual survey execution. While much has been said in the literature about the technical aspects of CATI, little attention has been paid to the more social aspects of introducing a CATI system to a research environment (Berry and O'Rourke 1988:458). These aspects were found to be a major concern in the PRL hiring, division of labour, training, supervision, and survey execution.

Project Management

The Laboratory's initial experience with CATI suggests that the person managing the project should have a general awareness of the basic microcomputing operating system, the statistical package the data is headed for, and a wordprocessor that produces ASCII files and routines for data cleaning, in addition to experience with survey methodology in an academic social science research setting. It is doubtful that the CATI system used could be successfully implemented in an academic research setting without personnel competent to merge new technologies with established patterns of doing survey research.

Fortunately, the PRL has staff well versed in survey methodology and also familiar with microcomputers and software. Therefore, the PRL decided to use its own personnel to implement the CATI system rather than to contract the computer programming out. A graduate student researcher familiar with the PRL survey process was also hired to operate in-house. This person was given responsibility for the software programing, questionnaire design and modification, training, some supervision, data processing, and preliminary statistical analysis. Permanent laboratory personnel looked after sampling, budgeting, hiring, and clerical tasks. The researcher managing the project was to consult with the Director and main research technologist on each major step in the survey process. Temporary staff were hired to do the supervision, interviewing, and coding. This particular division of labour was necessitated by the concurrent involvement of the PRL's permanent personnel in other projects. In other words, the implementation of the new CATI system in the planned survey project depended heavily on where one person was responsible for many of the survey's technical tasks, and data processing leaving little time for supervisory tasks such as monitoring interviewer performance and editing questionnaires from the field.

Training

A major consideration of incorporating the CATI system into a survey research setting is the skill level and training of the interviewers. Only one interviewer in the study had experience with CATI, though not the same CATI system as the one used by the PRL. Interviewers available from the PRL pool of interviewers had no experience with computers. An advertisement in the classified sections of the city's two major newspapers for interviewers with typing skills and computer experience mentioned as an asset brought inquiries from several people with a computing science background but no interviewing experience. Clearly at the time of the study, CATI was not used extensively by survey practitioners in the city of Edmonton.

During the initial stages of the survey, materials from previous telephone surveys, most notably the telephone interviewer's manual and handbook, had to be modified for the CATI system. In addition, materials were developed to explain the editing features of the CATI system. Interviewers were then given this material to digest before the pre-test training session. We suggest that the pretest should be conducted using the full CATI system rather than a pencil and paper techniques. Using this modified training strategy both the CATI system and questionnaire content problems can be dealt with simultaneously. After a very brief training session for the pretest many of the problems of introducing the CATI system to the PRL's experienced interviewers surfaced.

During the training for the pretest, interviewers were shown each item on the pretest questionnaire and instructed on the computer procedures associated with each item. After the brief training and introduction to personal computers, interviewers were then requested to do some trial interviews with a friend over the phone and enter the responses on the computer. Later they would begin interviewing the pretest sample. Severe problems

developed with this training strategy. The interviewers did not find the computers to be user-friendly and in some cases felt great apprehension about even touching computers. Unfortunately, the supervisors had not used the computer either. These problems seem to be common to installations considering CATI use for the first time and have been reported elsewhere (Spaeth 1987:22). The researcher in charge of the CATI system had to solve both the technical and social problems associated with first time computer use. It was clear that a different strategy had to be implemented in training the interviewers for the main study.

Some of the features of the pretest training were incorporated into the main training session. Once again a computerized overhead was used to demonstrate and discuss each question in the study. As in pretest interviewers did a few practise interviews on supervisors, the PRL researchers and friends. The resulting interviews were scrutinized carefully before continuing with the main survey respondents. However, the main training session was modified substantially to deal more effectively with both the social and technical problems anticipated for interviewers with the survey.

A major apprehension in using the computer for the first time is a fear of the unknown in terms of operating the computer and coping with a vast array of available software. To help allay some of these fears interviewers were told repeatedly that the computer for this particular application was basically a large typewriter and that they did not need wordprocessing, spread sheet, programming, or other computer-type skills to do this survey. The CATI software essentially turns the computer into a large interactive typewriter for the interviewer. Also useful was the provision of hands-on experience with computers and the CATI

system. Mock interviews were conducted in a large computer lab with the CATI questionnaire to be used. Interviewers could punch the keys appropriate to the response and practise some of the editing functions built into the questionnaire. This strategy had a two-fold benefit. First, the mock interviewing session gave interviewers a less intimidating introduction to the computer. Second, the new strategy also helped to familiarize the interviewer with the questionnaire. In the view of the PRL's experience, careful training can overcome the fears and challenges that the use of computers and the CATI system presents, so allowing, in the long run, more emphasis on the quality of the interview.

Interviewers

In the course of the survey, interviewers were able to overcome their initial discomfort with computers and the CATI system and produce quality interviews. Interviewers reported very few technical problems after the first week of interviewing. Some interviewers even expressed an interest in learning more about the computer. As mentioned previously, one potential outcome of using the CATI system is that it further centralizes the survey process. In the case of the PRL survey a couple of interviewers were permitted to interview from home using their own computers.

One problem that had to be adjusted during the survey was that after the interviewers became very comfortable with the questionnaire there was a tendency to race through the questionnaire. An interview that initially took around half an hour to complete could be completed under 15 minutes towards the end of the survey. Supervisors detected this problem early and instruct ed interviewers to slow down their pace in asking questions.

Supervisory Tasks

For the most part, traditional methods of manually scheduling calls, callbacks, keeping track of refusals and calculating response rates were employed in the mental health survey. Supervisors were responsible for keeping track of personnel and callsheets, computing daily interviewer and survey tallies as well as validating completed survey questionnaires. Many of these supervisory functions could be automated with a networked version of CATI. Because of the financial and computing resources available for the study a CATI system for stand-alone computers was purchased. Another consideration affecting a more automated supervisory role was that the supervisors were not familiar with the computer, a severe handicap in trying to implement automation of survey tasks. In the future, supervisors involved with the CATI system would need some wordprocessing skills and basic knowledge of the computer hardware and operating system.

Ordinarily, in a pen and paper telephone survey, the supervisors would also be very heavily involved in editing the questionnaires coming back from the field. In the PRL mental health survey they were not. A major bottleneck developed as a result in the simple editing of the completed questionnaires. While the CATI system can provide assistance to the interviewer on some questions, other questions, especially open-ended types are subject to a high degree of subjective interpretation. A separate computer was required at the time of interviewing to edit the questionnaires on-line. After each shift several interviews would have to be edited and the diskettes backed up. Ultimately, this involved over 800 interviews. The majority of the editing was done by the PRL researcher in charge of CATI, since the version of CATI used by the PRL requires completed interviews, fully edited and validated for merging into a database ready for cleaning and summary statistics.

A result of the awkward editing process in this survey was that the validation of surveys by the supervisors was held up and the majority of data processing, merging and coding of open-ended questions was started after the survey interviewing was completed. This is a serious limitation since validation and prompt editing can contribute substantially to overall survey quality by monitoring interviewer's collection of answers and, of course, serve as a deterrent to the "fudging" of interviews. Once again it is anticipated that many of the problems experienced during these phases of the survey would be lessened in the networked version of the CATI system where computer-trained supervisors can interact with the interviewers in the same computer environment.

Hardware

In retrospect, the PRL was not really set up to conduct a CATI survey given the hardware and personnel requirements. All computers at the PRL are in constant use for administrative, clerical and data processing on other projects. The computers used in the study had varying hardware configurations. Moreover, the design of the survey and implementation of the CATI system centralized the survey process in facilities which were already extremely busy. For the study during the day, only two interviewing stations could be in use and at night, after the PRL permanent staff had departed a maximum of four stations could be operable. In essence, the study had to be squeezed in around the day-to-day functioning of the PRL using machines not dedicated to CATI interviewing. Although the PRL staff was very accommodating on the initial CATI study, clearly alternative arrangements are going to have to be made in terms of space and facilities in future CATI projects undertaken by the PRL. A separate location with networked computers dedicated to CATI interviewing and supervising is one recommendation based on the initial CATI survey experience.

A number of problems in implementing the survey steps were anticipated and almost completely different problems were experienced. Interestingly, the PRL decided to run several paper hardcopies of the study questionnaire as a precautionary measure in case "the CATI thing" did not work. One of the hardest challenges in the planning process was keeping track of diskettes and CATI interviews in the same orderly fashion as the stacks of paper questionnaires coming in from the field in the traditional survey process. Call dispositions and interview scheduling still used pencil and paper techniques while the rest of the survey was handled on-line with diskettes, which were backed up after each interviewing shift. The result was two electronic questionnaires for every respondent. In total, there were 823 respondents in three different sub-samples and 64 diskettes to keep track of during the survey.

Across the study, very few technical problems were experienced. Only one interview had to be redone due to hardware failure. Minor problems experienced during the survey included disk drives making noise during the interview, some computer keyboards having an excessive click, and initially some interviewers would place the diskette improperly in the drive. However, by maintaining a continual backup of the interviewing diskettes technical problems that could result in the loss of several interviews were avoided or minimized.

Telephone headsets, however were a continual source of problem. They were used to allow the interviewer to have both hands free while typing. Although several headsets were tested, borrowed and even purchased during the duration of the study, no satisfactory solution was found. For one thing, different telephone lines (local, WATS and FX used to contact the three separate sub-samples in the survey) required different makes of headsets. By the end of the study some interviewers preferred to use a simple headrest for the telephone handset that comes with a phone. Newer models of telephone headsets may alleviate many of the problems experienced with headsets in future projects.

Software and Questionnaire

In the version of the CATI system used by the PRL it was found that all the question types could easily be accommodated. The questionnaire included skip patterns, ranges on question choices, variable text insertion, arithmetic checks for consistency between responses, open-ended questions and close-ended questions with an open-ended category for later coding. One time-saving feature of this CATI system was that the questionnaire could be produced on a wordprocessor and then imported into the program. The remaining programming task for the person constructing the CATI questionnaire was to link up instructions and questions from the wordprocessing file with the logic behind the questionnaire in terms of routing, ranges, and question choices. Introductory instructions and a statement asking for the respondent's cooperation were kept separate from the electronic questionnaire. We felt that in the absence of the networked version of CATI this modification to the introduction would improve the flow of the survey. In total, slightly over 100 questions were asked. The electronic questionnaire was very simple in terms of CATI capabilities and very easy to modify.

Data Processing

From the perspective of data cleaning, the CATI system can make a truly remarkable contribution to the survey process. A major advantage of CATI systems is the automatic generation of skip patterns. Another is the programmable restriction on acceptable ranges on question choices. What this means is that if a response is entered outside a permissible range the computer can

alerts the interviewer about his or her error. With automatic skipping or routing of the questionnaire interviewers can be prevented from asking the wrong question. These two programmable features are clear improvements over traditional pen and paper methods where questions and even entire pages of questions can be missed due to complex routing in the questionnaire. Costly callbacks to correct routing errors can be avoided. Besides contributing to reduced field editing of the questionnaire, we found that the resultant data set from a CATI setup required very little additional cleaning.

The process followed by the PRL was to wait until all interviewing and checking was essentially completed before cumulating all the interviews into one file in order to start coding the open-ended questions. An electronic coder was used where coders coded the responses on-line as the responses appeared on a computer monitor. The PRL coding staff found the electronic coder extremely fast compared to the traditional method of determining codes and placing them on IBM coding sheets ready for keypunching manually. However, the more positive social aspects of coding with several people working around a table were lost when a coder was faced with a computer screen in isolation.

After the open-ended questions were coded on-line, the close-ended and open-ended data files were merged and the data exported to SPSS/PC+. The data were cleaned using SPSS/PC+ DATA ENTRY II and then transported via ASCII portable files to the university mainframe for statistical analysis with SPSSx. Smaller runs were done on the microcomputer. The turnover time between completion of interviewing and the production of a machine readable data file was approximately one twelfth of the time required for traditional methods using keypunching to enter the data and extensive cleaning algorithms.

Cost-Efficiency

It is difficult to evaluate the costs of CATI versus traditional pencil and paper given the varying scope and nature of the telephone survey projects going through the PRL. On the one hand, a quick subjective assessment of CATI costs indicates that, considering the initial start up costs in terms of software and managing personnel the study might have been conducted on a cheaper basis using traditional pencil and paper techniques. On the other hand, a more favourable assessment of the cost-effectiveness of CATI might be forthcoming after several CATI studies are completed and the initial start up costs absorbed.

A rough comparison of the estimated budget and final expenses showed that while computing costs in the form of coding, cleaning, and keypunch are substantially reduced with CATI the costs of supervision in the survey escalated. The rise in supervision costs reflect problems inherent both in the editing process and division of labour. It is anticipated that with the addition of a networked CATI version that supervisors would spend less time completing survey tallies and sorting calls and more time on editing and validation during the survey process. Furthermore, using a networked CATI system the PRL person managing the project might spend less time supervising interviewers and editing questionnaires.

Overall Assessment

The CATI system is a useful addition and enhancement to survey projects that require telephone interviewing. The speed and efficiency that the electronic questionnaire contributes to the data processing is a major contribution to turnover time and cost. A major consideration in the integration of CATI into an academic research setting accustomed to using pencil and paper techniques is the degree of adjustment required. In the case of the PRL experience with CATI the transition was relatively smooth given the constrained computing resources and lack of computer expertise of supervisors and interviewers. Many of the obstacles to successful implementation of CATI were cleared during this initial run. While traditional pencil and paper methods are still used by the Population Research Laboratory for its main annual survey, the addition of networked computers and CATI version capable of automating many of the current supervisory tasks would further establish CATI in the Department as a preferable survey improvement. Other installations with CATI report the advantages of a more automated system including sample selection (Sharp and Palit 1988). According to a recent survey of 42 academic social science survey facilities in the United States and Canada the PRL is now one of two installations in Canada reporting the use of a CATI system (Spaeth 1990).

References

Berry, S. and D. O'Rourke.1988. "Administrative designs for centralized telephone survey centers: implications of the transition to CATI." Pp.457-474 in R. Groves, P. Biemer, L. Lysberg. J. Massey, W. Nicholls II and Joseph Waksberg (eds.), Telephone Survey Methodology. Toronto: John Wiley and Sons.

Currie, R. 1986. Selected Findings From the 1986 Winnipeg Area Study. Winnipeg Area Study Research Report No. 12. Department of Sociology, The University of Manitoba, Winnipeg, Manitoba.

Kinzel, Cliff. 1989. All Alberta Area Study 1989 Sampling Report. Edmonton Area Series Report No. 65. Population Research Laboratory, Department of Sociology, University of Alberta, Edmonton, Alberta.

Nicholls. 1988. "Computer-assisted telephone Interviewing: A general introduction." Pp. 377-385 in R. Groves, P. Biemer, L. Lysberg. J. Massey, W. Nicholls II and Joseph Waksberg (eds.) Telephone Survey Methodology. Toronto: John Wiley and Sons.

Prewitt, Kenneth. 1983."Management of survey organizations." Pp.123-144 in P. Rossi, J. Wright and A. Anderson (eds.) Handbook of Survey Research. New York: Academic Press.

Sharp, H. and C. Palit. 1988. "Sample administration with CATI: The Wisconsin Survey Research Laboratory's System." Journal of Official Statistics 4(4):401-413.

Spaeth, M. 1987. "Cati facilities at survey research organizations." Survey Research 18(3-4): 18-22.

Spaeth, M. 1990. "Cati facilities at academic survey research organizations." Survey Research 21(2): 11-14.

An important factor in the PRL's decision to use CATI for this project was a conference on state-of-the-art telephone survey methodology, including CATI, attended by the PRL's research technologist.

The PRL purchased the CI2 system for stand-alone computers developed by Sawtooth Software. In addition to the CI2 system the CI2 electronic coder was also purchased from Sawtooth Software.

The PRL Alberta Study (including the Edmonton Area Study) is conducted annually throughout the province of Alberta. Respondents in the City of Edmonton are interviewed face-to-face and the rest of the sample across the province is selected using random digit-dialing techniques and interviewed by telephone.

[1] Dave Odynak Researcher Population Research Laboratory Department of Sociology The University of Alberta Edmonton, Alberta and Cliff Kinzel Research Technologist Population Research Laboratory Department of Sociology The University of Alberta Edmonton, Alberta T6G 2H4 ph. (403)-492-2659

Evaluation and Appraisal of University Administrative Computing Datasets

by Mark Conrad [1]
Data Archivist
Pennsylvania State University

Introduction

In September of 1990 the Pennsylvania State University Archives in cooperation with Management Services, a division of the university responsible for administrative computing services, began a two year grant project funded by the National Historical Publications and Records Commission (NHPRC Grant #90-095). The objectives of this project are to appraise, preserve, and make available electronic records created, stored and used in the Management Services Division of Penn State University; to develop ongoing procedures for the appraisal of administrative computing data in the future; to develop protocols for the use of the data by institutional and outside researchers while enforcing restrictions on access for privacy and confidentiality purposes; and to provide recommendations based on the project for the preservation of archival data from on-line administrative database systems.

The project was divided into four phases. The first phase lasting two months was used for the orientation of the data archivist to the operations of the University Archives, the University Records Management Program, and Management Services. Phase two, the phase we are currently operating in, is used for the appraisal of datasets. This phase is scheduled to last 18 months. Phase three and four are scheduled to last the final four months of the two years. In phase three recommendations will be made for the identification and preservation of future archival datasets and protocols will be developed for the research use of the datasets. In phase four reports on the project will be prepared and circulated. In actuality, much of the work scheduled for phases three and four has already begun and is being carried out concurrently with the appraisal process.

This paper will focus on the second phase of the project. I will discuss the appraisal process as it is currently being carried out, difficulties encountered and lessons learned to date.

The Appraisal Process

For the purposes of this project the appraisal process will be confined to a finite number of datasets from the university's administrative computing mainframe. We will not be examining electronic records from other mainframe, mini or microcomputer systems.

Rather, the appraisal process will be limited to some 3,000 datasets recorded on "history" tapes. History files, in Management Services' parlance, are usually copies of master files at a particular point in time (often the end of a semester or academic year). The datasets would be very difficult to recreate if they were destroyed because they are copied from files that are constantly updated. These files are kept for possible reuse [2].

The datasets date from the late 1960's to the present. These datasets contain information for fourteen areas of administrative responsibility at the university. The areas are: Accounting, Payroll, Bursar, Student Aid, Agriculture, Planning and Analysis, Budget and Resource Analysis, Management and Systems Engineering, Admissions, Registrar, Testing Services, Development, Graduate School, and Physical Education.

Each area has a data steward who "develops the coding structure of the data, insures the data's accuracy, determines the frequency of updating, and establishes data use and protection requirements."[3] The data steward is usually a senior administrator, such as the Registrar, or that persons' designate.

The appraisal process begins by selecting a data steward's area of responsibility. The data archivist must have the written permission of the data steward to examine the datasets under his/her control. In some cases datasets are jointly "owned" by more than one data steward. In such instances the disposition of the datasets must be discussed with all interested parties. In starting the appraisal process I chose data stewards that had only a few history tapes to test the procedures developed for the appraisal process and the database system to track appraisal information.

Having chosen an area of responsibility, the next step is to identify those datasets that belong to the data steward from the (Management Services) Tape Library listing of the history files. Once identified the datasets are grouped by common dataset name. This is done because datasets with the same name usually contain the same types of data and can be initially evaluated as a group.

The next step is to locate as much information about the datasets as possible. I am to find the procedure and

program that created or used the information, any documentation, a file description, a record description, record counts, or any samples of input or output. Records Management Program retention schedules are checked to see if similar records in another format have already been scheduled. If similar records have been scheduled for destruction and the datasets do not have some additional value by virtue of their being in electronic format and thus more manipulable, the datasets may be recommended for destruction. (Junk is junk no matter what the format.)

While the search is on for information about the datasets, some of the tapes are read and printouts are made of a sample of records from each file. This serves several purposes. Firstly, it verifies whether or not the tapes are still readable. Some of these tapes have been in storage for a very long time under less than ideal environmental conditions. Secondly, the dump can be used for comparing what's actually on the tape with what the documentation says should be on the tape. If the two don't match, other documentation must be located or the file may be recommended for disposal. A file is of no value if a determination cannot be made as to where one field ends and the next begins or as to what a particular value in a field indicates.

Having gathered as much information about the dataset as possible, the next step is to interview the data steward and/or a contact designated by the steward about the datasets under his/her control. A standard list of questions has been developed to help the data archivist gather all the information necessary to make an informed appraisal decision.

Those questions are:

Do you have documentation for these files?

Do you have samples of input and output for these files?

Where did the data come from?

What was it used for?

Is it still being used?

Is it updated?

How often?

Are the records maintained in another format?

Has the other format been scheduled for retention or disposal?

Are there any requirements for the retention of this data that you are aware of?

Are there any restrictions on the use of this data that you are aware of?

At this point the data archivist should have enough information to begin making the appraisal decision. The decision-making process is not that different from the process for more traditional records. It is certainly not very different from the process used by archivists working in the electronic records programs in government archives.

Does the dataset have legal, evidential, or informational value?

Is this dataset unique?

Is it the most desirable format for keeping the information?

Is the data hardware/software independent?

If the answer to all of these questions is yes, all datasets that share the common dataset name and structure will be recommended for accessioning by the Archives. Retention schedules are developed for all datasets, regardless of their status, in concert with the data steward, Management Services, the Records Management Program Staff and the University Archives/Records Management Advisory Committee.

Once the decision has been made that a group of datasets are archival, each dataset is read and a data dump is obtained. As with the sample of datasets read previously this is to verify that each dataset is readable and the data is valid. File structures and record descriptions change over time so the data archivist must insure that data from each dataset is adequately documented so that a researcher or the archives staff can use it. Assuming the datasets are readable and understandable, copies are made of each dataset and the relevant documentation. The datasets are accessioned by the University Archives and the data archivist turns his attention to the next set of files.

Problems Encountered

You may have already gathered that one of the biggest problems has been locating adequate documentation for many of the datasets. The record description for many files seems to change on a regular basis. Often the documentation is not updated to reflect these changes or conversely when the documentation is updated previous versions are discarded despite the fact that files still exist that were created using the previous documentation. It is not unusual to find a number of files with different file

structures, the same name, and one set of documentation that may not match any of the files. In talking with other archivists working with electronic records, I have been assured that Penn State is not alone in this predicament.

Management Services is currently exploring the possibility of recording documentation for a dataset directly onto the first label of the tape a dataset is recorded on. As long as the documentation is copied along with the dataset whenever the dataset is transferred to new media, the proper documentation should be available for the life of the dataset.

Another related problem occurs when trying to appraise an older dataset. Often there are no employees still working in the office that used the file who remember what the file was used for. Sometimes the office itself no longer exists! The turnover on a university campus and the restructuring of administrative units can make it difficult to find someone who can tell you how a file was originally used or what dataset replaced the one you are evaluating. The only solution to this problem is to carry out the appraisal early in the life cycle of a dataset.

Lessons Learned

One of the most important lessons we have learned is that the shorter the time lapse is between the creation of a dataset and its appraisal the easier it is to identify archival datasets and insure their preservation. The archivist can interview all the players involved in the creation of the records to better understand why the records were created and under what circumstances. Documentation can be evaluated to insure it adequately explains the data so that it will be useful to future researchers. Datasets that are identified as archival can be marked for special handling to insure the data will still be readable 10, 25, or 100 years from now.

Another lesson we have learned is that the cooperation of the administrative computing center is essential to the success of an electronic records program. The archivist needs to understand how data is manipulated at the center to meet the informational needs of the institution. The administrative computing center personnel must have an appreciation of the potential value of the data beyond the purposes for which it was originally created. Any archives considering implementation of an electronic records program would be well advised to begin building relationships with their institution's administrative computing center(s) now.

The most important lesson we have learned is that more records are being stored in electronic format all the time. If we do not identify and preserve the archival datasets a large portion of our institutional memory will be lost. At Penn State we have begun the process of insuring these valuable records will be preserved, we encourage other institutions to join us, and we are happy to share information about our project.

Footnotes

[2] Pennsylvania State University. Management Services Division. Standards and Procedures Manuals (on-line manual).

[3] Pennsylvania State University. Administrative Policy, AD-23.

References

Pennsylvania State University, Management Services Standards and Procedures Manual #3, Chapter 5, Section 1. (on-line manual)

Pennsylvania State University Administrative Policy, AD-23, p. 1.

[1] Paper Presented at IASSIST 91, 15 May 1991, Edmonton, Alberta, Canada.

Remote Access to Local Data Collections: A UC Davis Experiment

by Jean Slemmons Stratford [1]
Head Librarian, Institute of Governmental Affairs Library, University of California, Davis

INTRODUCTION

The East Asian Business and Development (EABAD) Research Archive (Institute of Governmental Affairs, University of California, Davis) is currently in the process of implementing a system that will facilitate remote access to its local data collections. This paper provides background on the Research Archive and discusses the development, design, and organization of its public access system, *Asia Online*.

BACKGROUND

The EABAD Research Archive houses printed and machine readable information on East and Southeast Asian corporations and economies. Established in 1986, the EABAD Archive is the only facility in the United States that systematically collects materials supporting research on the firms and inter-firm networks of Asia. [Background on the Archive and the research projects it serves was provided in a previous paper "East Asian Business and Development Research Archive: A Unique Data Resource" *IASSIST Quarterly* 14 (Summer 1990) :3-8. In addition to its printed collections, the Archive is one of only two academic facilities in the world that acquires and develops computerized databases supporting the study of Asian corporations and economies. The other is the University of Hong Kong's Mong Kwok Ping Management Data Bank.

In 1990, the Research Archive was awarded an infrastructure grant by the President's Office of the University of California. One of the primary components of that grant project is the establishment of an online computer system to facilitate inter-campus access to the resources (both print and machine readable) of the EABAD Archive. The grant provided a one year award of $75,000 to establish the online system in the hopes of enhancing scholarly cooperation and access to resources supporting research on Asia on the 9 UC campuses. The system presents some options for extending access to highly specialized, yet topical, research data. It should also provide an opportunity to assess the feasibility of remote access to local data collections and to gain practical experience that will allow us to identify the strengths and limits of our chosen remote access system.

SYSTEM SPECIFICATIONS

From its inception, the system was intended to be accessible both via direct Internet connection and via the UC systemwide online library catalog MELVYL. Mounting the system on the Internet provides scholars in the UC system and in many locations worldwide with direct, interactive access to the system. A linkage to MELVYL provides a high degree of visibility for *Asia Online* within the UC system, while its specialized bibliographic files complement and extend the usefulness of MELVYL. The requirements for these connections guided the selection of both system hardware and software.

The system is running on a 33 Mhz 80386 PC with 8 MB of RAM and a 338 MB hard disk drive. For security reasons, it was deemed best to mount the system on a dedicated machine, and a large, fast PC provided the most cost effective platform for this purpose. In addition, the EABAD research program has been fortunate enough to develop a donor relationship with the Taiwan-based computer manufacturer Acer. In 1989, Acer donated 3 20Mhz 80386 PCs and a file server with 326MB hard drive and Novell Netware. This hardware serves as the core of our in-house LAN. In the spring of 1990, Acer America Corporation donated the hardware for the system.

The operating system of choice for the project was UNIX. There are several reasons for this. First, the UNIX implementation of the TCP/IP protocol facilitates easy interactive access to both the Internet and MELVYL. UNIX's open architecture also makes program development relatively fast and easy. The Santa Cruz Operation, the manufacturer of the top selling UNIX for the PC environment, donated its operating system to the project. In addition, an SQL DBMS, Integra, that incorporates links to the C programming language was acquired for the database functions of the system. Using these products a user interface is being developed that supports "dumb terminal" access. The system software takes a "plain vanilla" approach in order to support the broadest possible base of users and to interface easily with MELVYL.

SYSTEM DESIGN AND ORGANIZATION

As initially conceived, the online system will have 3 basic components. These are outlined briefly below.

First, the system will feature a public message service (or "news" system) to facilitate access to information about Asia-related research activities. Types of information on this service may include the following:

announcements of relevant events (lectures, conferences, seminars, etc.),

information on funding sources for Asia-related research,

general information of interest to Asia scholars in the UC system,

publications lists from the numerous institutes and research programs on the campuses,

research related information including abstracts of UCOP Pacific Rim projects or other research programs,

online versions of newsletters from such groups,

inquiries from scholars,

listings of appropriate research resources on UC campuses (computer files, library resources, grant programs, etc.).

A second component of the system will be online access to select public domain statistical data held by the Research Archive. These files will be mounted in a standard microcomputer file format (such as dBase compatible or Lotus worksheet), along with documentation files and will be available for downloading and local use by scholars.

Finally, we will incorporate bibliographic files in the EABAD system. The first will be a database of our own holdings (both print and machine readable). We also intend to mount the United Nation's *Asian Bibliography*. *Asian Bibliography* is a semi-annual publication of the UN's Economic and Social Commission on Asia and the Pacific that provides citations to a wide range of literature in the social sciences.

At present, the prototype system has just come up for public access. It supports all of the functions outlined above except the bibliographic and statistical files which are expected by Fall of 1991.

Once the system is fully established, the final activity for the project will be an evaluation. This evaluation will be based a variety of information. A primary concern will be the extent to which the online system has strengthened the activities of the Research Archive. This may include such subtle and non-quantitative measures as the extent to which it has fostered ongoing research, new collaborations, or greater awareness of the research activity of EABAD. A second method of evaluation will be based upon data gathered from online users. This will include both an examination of the volume of usage, the distribution of users from among the campuses and academic disciplines, the intensity of usage by individual users, and any feedback received from online users, and the extent to which interlibrary lending has increased due to the online database of EABAD holdings.

Presented at the IASSIST 91 Conference held in Edmonton, Alberta, Canada, May 14-17, 1991.

History and Statistical Analysis: A case study

by Patricia E. Prestwich[1]
Department of History
University of Alberta

In my historical work, I am an occasional, or perhaps even accidental, user of computerized statistical analysis. My level of competence in this area can be best conveyed by the fact that I first met my colleague, Chuck Humphrey, of the University of Alberta, when I walked into his office and told him that I had copied the records of 14,000 French mental patients. I then asked whether he thought that I could analyze them using index cards. I was fortunate to find an expert who understood what I was trying to do with my data and who could make the computer work for me.

Historians have long been reluctant to engage in extensive statistical analysis, which they often dismiss as "number-crunching." In part, their reluctance stems from a genuine fear of obliterating the particular and the personal—aspects that, for many of us, are an essential part of history. Obviously, however, this reluctance also stems from ignorance or fear of the technology and methodology.

The field in which I am now working—the social history of medicine or, specifically, the social history of madness—illustrates how slowly historians can turn to computerized statistical data. This is a relatively new field and until seven or eight years ago, most people in the field concentrated on the analysis of historical documents, particularly the writings of doctors, using many of the theories about power inspired by Foucault and sociologists. A primary interest has been the nineteenth century psychiatrist hospital, or asylum,or "madhouse". It has been seen as the symbol of social control, of the ways in which the bourgeoisie in general and psychiatrists (or "mad doctors") in particular deflected any challenges to their power by labelling it as deviation.

But, as a number of historians in different countries began to point out, much was being theorized about the asylum without any detailed evidence of how it functioned or whom it supposedly controlled. For the past few years, a small number of studies have emerged which look at asylum records and try to understand the complex functioning of this institution. These studies , although few in number, have already begun to challenge many of the predominant theories about the asylum and about nineteenth century psychiatric medicine. Most of these studies contain some statistical analysis, although even historians of the asylum are still cautious in this respect.

My own research is the study of a Parisian asylum, Sainte- Anne, from its opening in 1867 as the first of the new model asylums, until the end of the First World War. Sainte-Anne may not be a typical asylum—although no one is sure now what a typical nineteenth century asylum was. Like most public asylums in the nineteenth century, it was for the poor, in this case the working class and petty bourgeois of Paris. Sainte-Anne was, however, the only Parisian asylum that was not in the suburbs, but the city itself—an important factor in considering the relations between families, the asylum and the psychiatrists. It was also the teaching hospital for the Faculty of Medicine of the University of Paris and its doctors were among the most eminent in France.

The nineteenth century asylum generated masses of printed statistics— in fact the main occupation of nineteenth century medical institutions seems to have been the compilation of statistics. This is not only a reflection of their institutional character but of the fact that by the end of the nineteenth century doctors seemed to be more interested in the diagnosis, or rather the classification, of mental illness than in its treatment. Data on mental patients became an important means of both refining and justifying their classifications.

But, of course, much of the published statistical material is not useful today because we ask different questions. To give some specific examples, the asylum recorded and printed extensive statistics on the occupations, marital status, age, sex, and diagnoses of their patients but always in separate charts, so that it is difficult to make any correlations. (For example, we know how many single women were interned, and how many employees, but not how many single women employees.) They recorded the length of stay of those admitted for the first time (probably with a view to giving a rosier picture of cure rates) but not of those who had been readmitted, although readmissions constituted a significant proportion of their patients. In the printed statistics, there is no correlation between diagnosis and length of stay, or between length of stay and result of treatment (i.e. death, transfer or release). So, for example, it is impossible to tell from the printed records whether a male depressive

would stay as long as a male alcoholic or a female depressive and what chances each had of release. Thus, while the printed material is sometimes useful for verification, it was essential for me to compile my data from the original records.

These records are the Registres de la loi, the legal register that must be retained permanently for every patient admitted to psychiatric hospital in France. They are highly confidential documents and even today are not computerized because the French have very strict legislation about privacy of 'information. (Today, a clerk enters the details by hand; in the nineteenth century it was often mental patients who did this work.) The Registers give the basic demographic data on each patient— age, occupation, marital status— as well as date of entry, date of exit, legal status, and result of treatment (ie death, transfer or discharge.) There are also three diagnoses for each patient: an admitting diagnosis, a diagnosis after 24 hours and a diagnosis after 2 weeks. Usually, the diagnoses are by different doctors. The records often contain incidental information on the circumstances under which the patient was interned (e.g. as a result of a suicide attempt, family violence, or strange behaviour) and sometimes some observations of the patient's behaviour while interned.

I have selected the Registers only for Sainte-Anne itself. The hospital also had an Admissions Bureau which saw almost every patient that was interned in the Paris region. The patients came through this Bureau and were sent on to the various Parisian asylums. Approximately 3000 patients per year passed through the Bureau of Admissions and their records are intact, including many of their medical files. To collect the data, would, however, be an immense project that could only be undertaken by team effort. My data come from the patients that were transferred from the Admissions Bureau to Sainte-Anne itself. The asylum was built for 500 patients, but by the 1890s usually held about 1000 patients. I have transcribed the registers for every second year from 1867- 1927, for a total of 14,000 patient records. This sample is considerably larger than in comparable historical studies of public asylums, which usually select only certain years. I collected such a large sample in part to deflect criticism that my sample would be unrepresentative, but also because I felt that with a larger sample I could begin to ask certain questions about internment patterns that could not be asked with a smaller sample. Even now, I have certain problems; for example, I have only 238 cases of senility for the period 1873-1913 and so for some of the detailed analysis, my sample is extremely small.

Of course, even on the basis of selecting every second year, my sample is not complete, because certain registers could not be found. The registers are stored, in a very disorganized fashion, in a basement room, lit by a 40 watt bulb and covered in dust and rat poison (The basements of Sainte-Anne connect with the catacombs of Paris.) With the help of a hospital worker, or occasionally, a patient, I had to haul these large registers up from the basement. I simply did not find all the years that I wanted. Or, as often happened, since one year would be spread over several registers, I would find only part of a year. The registers were also difficult to read, because, apart from the dust and yellowing paper, the ink had faded and the handwriting was not always decipherable.

Although these registers offer some very difficult problems of interpretation, they are an important source for the type of social history that I am trying to write. My goal is to write a book on the asylum as a social institution, i.e. as part of a specific historical community. I want to understand what roles this medical institution played in the lives of families, patients, nurses, and doctors.I want to understand what power these different groups had and how they interacted. The statistical data is merely the beginning of my analysis. The data, in some cases, will give me specific answers, but in most cases, it will direct me to the nonstatistical literature.

For example, analysis of the statistical data is helpful simply to clear away some of the myths about the nineteenth century asylum and to establish who got interned, for what diagnosis and for how long. Social historians of medicine, who have read only the qualitative material, have postulated that the asylum was the dumping ground for the " inconvenient" in society, those who simply did not fit into the developing industrial society. Patients in public asylums were certainly not middle-class, but as the analysis of occupations at Sainte-Anne shows, neither were they the dregs of society. There were very few labelled as "vagabonds"(1.5%) and in fact, most gave their occupations as skilled workers (carpenters, seamstresses, etc.) or as employees.(43% and 16% respectively, but the figure is probably higher if one counts part of the 17% who were women listed as " no occupation and who are usually the wives of skilled workers or employees.) The proportion of unskilled workers, such as day labourers or domestic servants, in my data was only 14%. (Again if wives are counted, it might be higher.)

It is also clear that, once inside the asylum doors, patients were not necessary doomed to perpetual confinement. After about 1860, there was a great deal of political and public hostility toward asylums, which were labelled as "modern Bastilles", where people languished in unjust internment. Although doctors certainly had extensive legal powers, an analysis of the length of stay of patients over 40 or 50 years paints a more complicated picture. At Sainte-Anne, in the period up to the First World War, about 45% of all patients were released, 30% died and

25% were transferred. The length of stay for those who were released is shorter than one would expect.

Release:	25%	50%	75%
all :	40 days	94	220
cut to : 800 days :	36	78	162

Of course, these statistics can only be interpreted by relating them to the diagnoses. For example, the 30% death rate, which was higher for men than for women, is directly related to the high number of male patients interned for general paralysis, the third and fatal stage of syphilis . (General paralysis made up 22% of male internments. Eighty-Seven % of GP cases were men and the death rate at the asylum itself was about 75%.

The analysis of the data is useful simply to give some idea of how patients were diagnosed and, although my analysis of this aspect is not finished, there seem to be fairly discrete diagnosis, with not too much overlap. The most common diagnoses were general paralysis , alcoholism, depression, persecution and old age in various forms. It is revealing to compare what doctors faced in the asylums— quite often what they would label " banal" or "uninteresting" problems— and what they discussed in their medical literature, which was usually the unusual, exotic or, as they said the " interesting".

The question of what interested doctors can be approached in another way through the data, for I have records not only from the asylum itself, but from the teaching clinic at Sainte-Anne. By comparing the patterns of diagnosis of the asylum and the clinic, I hope to make some deductions about what interested doctors and how comprehensive an education medical students received.

Aside from giving certain basic information about who was interned and why, the data can also begin the process of answering some of the questions about the role of families in the whole process of internment. One of the important aspects of the data is that admissions were divided into two types. The first was placement officiel (PO)—a legal internment which involved police action. Usually the person was taken to the local police station and then to the police dispensary, where a police doctor made the final decision as to whether the person would be sent to the Bureau of Admissions at Sainte-Anne. But by the 1880s, there was a second type of admission, the placement volontaire (PV), which allowed families and even friends to intern someone without going through the police, although this involved paying the internment expenses in most cases.

The PV admissions will give some insights into family behaviour, that is, what behaviour was considered so unacceptable or intolerable as to lead to internment and, conversely, under what conditions would families request the release of patients. This is not to imply, of course, that family decisions were not involved in placement legal. It is clear from the records that a number of families, presumably the poorer ones, would simply call in the police to deal with an intolerable family situation, such as an alcoholic father or a senile elderly relative. But the PV admissions give much clearer evidence of the family's role because they usually indicate who interned the patient (a mother, spouse, friend, etc.) . Also, because a patient interned "voluntarily" could be released at the insistence of a family member, even if the doctor objected, these files give some insights into the complex relationship between doctors and families.

One good example of family power comes from an examination of data on patients who were transferred. Transfer of patients from Sainte-Anne to more distant asylums became increasingly necessary as the asylum became overcrowded in the latter part of the nineteenth century. Transfers were strongly resisted, both by patients and families, because it usually meant transfer to poorer care and at a distance that made family intervention impossible. My analysis of length of stay shows that PV patients stayed considerably longer (i.e., in terms of years) than PO patients before they were transferred and that, significantly, this pattern was true for both men and women. I would argue that here is a clear indication of effective family influence.

A third aspect that emerges from the analysis of the data is the gendered nature of the asylum. Although feminist historians have speculated a great deal about the tendency to label women as mad if they did not conform to societal norms, there has been relatively little analysis of the asylum from the point of view gender. Again, statistical analysis is useful to clear away some myths. Women, for example, were not interned more frequently than men, nor did they have a lower release rate. But, they did stay longer and consequently, they had a higher rate of transfer. These differences are clearly related to different patterns of diagnosis. Women and men, on the whole, were diagnosed differently. The clearest example is between alcoholism and depression. Nearly 30% of the men, but only 10 percent of the women were diagnosed as alcoholic, whereas approximately 30% of the women were diagnosed as depressive, and only 10 % of the men. Men and women therefore had different experiences in the asylum. Why women were labelled as depressive and men as alcoholic is a question that cannot be answered by the statistical data, of course, but can only be explored by examining more traditional written sources.

This is my first foray into this type of analysis and I clearly have much still to learn. (Although I now admit

the superiority of the computer over index cards!) I wish that I had had some idea of the possibilities of computer analysis before I began to collect the data, but that was impossible. I obtained access to these records purely by chance; I recognized the their richness in terms of social history, but I simply had to trust that I would eventually find the right people and the right techniques to help me use the data. Whether I will ever use this type of analysis again will depend on the research project. My real problem now is to integrate this statistical analysis into a broader, more traditional narrative and to convey this analysis effectively to my audience of historians, who for the most part still skip the statistical sections in any book.

[1] Paper Presented to IASSIST Conference, May 17, 1991, Edmonton Alberta.

The U.S. Public Use Census Microdata Files as a Source for the Study of Long-Term Social Change

by Steven Ruggles [1]
Department of History
University of Minnesota

The United States public use microdata samples are machine-readable hierarchical files consisting of individual-level and household-level records drawn from the federal decennial censuses. Samples covering nine census years between 1880 and 1990 are currently available or in preparation. Taken together, these microdata comprise the richest source of quantitative information on long-term changes in the American population. Because these samples were created at different times by different investigators, however, they have incompatible documentation and a wide variety of record layouts and coding schemes. These differences among the samples inhibit their use as a time-series.

At the Social History Research Laboratory of the University of Minnesota, we are planning to convert the series of public use samples into a single coherent form. The success of this project will depend on the usefulness of the data series to a broad range of social scientists. This essay describes the history of the public use samples and some of their potential applications for time-series analysis, in the hope of stimulating interest and suggestions at an early stage of our work.

Background

Social scientists have increasingly recognized the need to study society as a process. If we confine our analyses to the state of society at a single moment, we cannot hope to understand the sources of social change. Sociologists, economists and demographers have developed a variety of quantitative data sources to study social change, including retrospective surveys, repetitions of early social surveys, and longitudinal surveys. Although such data sources are essential, they are usually limited to the analysis of changes during the past thirty years. The study of longer term change — over the past 100 or 150 years — has been sharply constrained by the limited availability of consistent data series. Analysts of nineteenth-century society have often turned to institutional and bureaucratic records, such as those generated by churches and the military, but these sources are typically available only for the distant past and they are limited to the study of specific population subgroups.

The decennial census is the most consistent general source of information about the American population over the past two centuries. Quantitative studies of long-term social change have always relied on the published tabulations of the census, but these data have substantial limitations. In each period, the topics addressed by census publications have focussed on contemporary concerns, and these concerns have shifted dramatically over the past century. For example, the early twentieth century census volumes include a wealth of data on immigrants, but virtually nothing on family composition. Moreover, the high costs of tabulation before the introduction of modern data processing equipment meant that few cross-classifications of census data were possible, and much of the information collected by the census was never tabulated at all. Even for recent census years, the published census volumes have significant limitations for the study of social change. Despite the dramatic increase in the quantity of published census data in recent years, the census Bureau cannot anticipate all the questions social scientists want to ask.

The Census Bureau has addressed these problems by producing individual-level public use samples of the census (U.S. Bureau of the Census 1972, 1973, 1982a, 1989). The first public use sample was created as a byproduct of the 1960 census (U.S. Bureau of the Census 1954). In an effort to meet the needs of scholars who needed specialized tabulations, the Census Bureau created a 1 in 1000 extract of the basic data tapes they had used to create tabulations for the published census volumes. To preserve confidentiality, the Census Bureau removed names, addresses, and other potentially identifying information.

The 1960 public use sample was an immediate success. Not only did it allow researchers to make tabulations tailored to their specific research questions, but it also allowed them to apply new methods to the analysis of census data, especially multivariate techniques. But the sample did have two significant limitations. First, the sample size was relatively small. The 1 in 1000 sample density yielded about 180,000 person records. Given the modest capacity of computers in 1964, this was a lot of cases, but as researchers began to use the sample for detailed analysis of small population subgroups, its limitations became apparent. Second, the 1960 public use sample provided highly limited geographic information. In its zeal to preserve confidentiality, the Census Bureau stripped off all information on places below the

state level. This meant, for example, that it was impossible to extract a subsample of the New York City population.

Both of these problems were addressed by the 1970 public use samples. The 1 in 1000 density of the 1950 sample was increased dramatically; the Census Bureau provided six independent public use samples for 1970, each of which had a 1 in 100 density. Users who required an exceptionally large number of caaes could combine the samples to obtain a six percent density, or about 12 million person records. In addition, the 1970 samples provided a variety of alternate geographic codes, although the Census Bureau still did not identify any places of less than 250,000 population.

In conjunction with the 1970 public use samples, the Census Bureau released a new version of the 1960 public use sample. They enlarged the sample density from 1 in 1000 to 1 in 100, and at the same time reorganized the coding schemes and record layouts to be compatible with the samples from 1970. This compatibility made it relatively easy for investigators to pool data from 1950 and 1970, and thus incorporate change into their analyses.

By the late 1970s, the public use samples had become one of the essential tools of American social scientists. It was in this climate that two separate teams of researchers independently came up with the idea of creating historical public use samples for earlier census years. Samuel Preston directed projects at the University of Washington and the University of Pennsylvania to produce a 1-in-750 sample of the 1900 census and a 1-in-250 sample of the 1910 census (Graham, 1980; Strong et. al., 1989). Meanwhile, Halliman Winsborough and a group of others at the University of Wisconsin and the Census Bureau created 1 in 100 samples for the censuses of 1940 and 1950 (U.S. Bureau of the Census 1984a, 1984b).

A fifth historical public use sample is now underway. At the University of Minnesota, we are creating a 1 in 100 sample of the 1880 census. That project is about half done, and a preliminary 1 in 1000 subsample is already available (Ruggles and Menard, 1990; Social History Research Laboratory, 1990). In addition, we have applied for funds to create a public use sample of the 1920 census; if that project is funded, the 1920 sample will be complete by 1997.

In the meantime, the Census Bureau has released public use samples for the 1980 census, and has scheduled a 1993 release date for samples of the 1990 census (U.S. Bureau of the Census, 1982a, 1989). These samples include greater geographic and subject content detail than either the 1960 or 1970 public use samples.

What all this means is that we can anticipate a series of public use microdata samples of the u.s. census covering the years 1880, 1900, 1910, 1920, 1940, 1950, 1950, 1970, 1980 and 1990. This data series will constitute a resource of unprecedented power for the study of long-term social change. The availability of the historical census files is especially important, because few national microdata files of any sort exist for the period before 1960. Furthermore, as one goes farther back in time the published tabulations of the census become increasingly sketchy and the problems of comparability increase.

Table 1 summarizes the availability of variables for each of the census years currently available or in preparation. Eleven basic questions were asked in all census years, and twenty-two inquiries are available for at least seven of the nine census years. There are a significant number of variables omitted from Table 1 that are available in only one or two census years. Note that in addition to the differences in available variables across census years, there are also multiple versions of the samples for recent years that incorporate slightly differing variables. A detailed discussion of comparability problems can be found in Ruggles (1991).

Applications of the Public Use Microdata series

The range of potential topics that can be addressed with these data is far too great to describe within the page limitations of this paper. The following paragraphs are intended only to suggest some of the most obvious topics of investigation.

1) Household Composition. American living arrangements have been radically transformed since the late nineteenth century. In 1880, for example, 77 percent of the elderly lived with their children or with extended kin, compared with 24 percent in 1980. The frequency of primary individuals has increased about eight-fold, and residence as secondary individuals or extended kin has dropped almost as dramatically. These changes began shortly after the turn of the century, and accelerated after 1940 (Ruggles 198B; Ruggles and King, forthcoming).

We are only beginning to understand the dimensions of change in family structure over the past century, and the analysis of the determinants of that transformation has yet to be seriously undertaken. The public use samples are the only detailed national source of information about changing living arrangements in the nineteenth century and first half of the twentieth century. All the public use samples provide sufficient information to construct fully compatible and highly detailed measures of household composition and family interrelationships.

2) Fertility. Between 1850 and 1940 the total fertility rate for White Americans declined from about 5.4 to 2.2 (Coale and zelnik 1963:36). Research on early fertility

trends in America has relied for the most part on child-woman ratios (Forster and Tucker 1972; Yasuba 1953) and backward projections of age distributions in the published census volumes (coale and zelnick 1953; McClellan and Zeckhauser 1982). Neither of these techniques allows close analysis of marital fertility or fertility differentials. Analyses of fertility using own-child techniques were among the earliest and most fruitful multi-sample studies carried out with the two original public use samples produced by the Census Bureau (e.g. Rindfuss and sweet 1977). The public use microdata series will permit study of differential marital fertility patterns over the period of greatest fertility decline, comparing characteristics such as race, occupational class, region, literacy, size of locality, family structure, and a wide variety of other variables. The richness of these data will greatly enhance our ability to analyze the determinants of early fertility decline in a developed country, and this may in turn lend insight into the onset of fertility control in developing countries.

3) Life Course Analysis. Long-term changes in the timing of major life-course transitions — such as leaving school, leaving home, starting work, marrying, and establishing a separate household — have been studied using both cross-sectional data (Modell, Furstenberg, and Hershberg 1975) and retrospective survey data (Hogan 1981). Both approaches reveal that American society has become more age-graded during the twentieth century: people tend to pass through the major transitions to adulthood at increasingly prescribed ages and in an increasingly prescribed sequence. Recently, Stevens (1991) suggested that the heterogeneity of the early decades of the twentieth century was a short-term phenomenon brought about by rapid urbanization and immigration from Southern and Eastern Europe. The public use microdata series will provide the opportunity to test this hypothesis through cohort analysis of both the timing of change and of differences among subpopulations.

4) Household Economy and Female Labor Force Participation. Much of the research on late nineteenth and early twentieth century social structure has focussed on patterns of employment within the household. Some investigators see a fundamental transformation of the household economy with the rise of wage labor; others point to the continued strength of preindustrial modes of informal family labor (Katz et. al. 1982; Anderson 1971; Barron 1984). Since the existing studies are based on small local samples of census data, regional variation may explain much of difference in interpretation. The hierarchical organization of the proposed census series is well suited to study of the household economy.

Female labor force participation is a closely related and equally controversial issue (Bose 1987; conk 1981; Folbre and Abel 1989: Goldin 1980, 1983; Openheimer 1970; Jaffe 1955). Changes in census definitions of employment and labor force participation have complicated such analysis. The public use microdata series will allow researchers to minimize the effects of such changes, since labor force participation can be allocated according to the procedures proposed by Abel and Folbre (1990); such adjustments are impossible with aggregate data. Analysis of the determinants of female labor force participation and child labor during the late nineteenth and twentieth centuries should prove especially revealing.

5) Ethnicity and Immigration. The questions on nativity in the public use samples makes them a rich lode of information for immigration historians. Throughout the period 1880-1970 the census asked about parental birthplaces as well as the respondent's birthplace. Most of the census years also provide information on mother tongue and year of immigration. This makes it possible to analyze patterns of acculturation for a wide variety of cultural groups. Understanding the varied experience of immigrants in the late nineteenth and early twentieth centuries has taken on a special relevance in light of the recent resurgence of immigration.

These topics are intended only as representative examples of the sort of research that can be carried out with the public use microdata series. Other key areas of investigation include the transformation of industrial and occupational structure, urbanization, internal migration, nuptiality, and education.

The large size of the public use samples increases their versatility by permitting analysis of small population subgroups. Consider, for example, some of the topics addressed by Minnesota graduate students using the historical public use samples:

- the professionalization of nursing
- American Indian fertility patterns
- race differentials in the living arrangements of the elderly
- labor force composition in Minneapolis and St. Paul
- the adaptation of scandinavian immigrants
- changes in the gender composition of clerical workers
- the household structure of early black migrants to

Northern cities

- Italian immigration to the Southern U.S.

- living arrangements of parentless children

These research topics could not be pursued using a general social survey of the scale ordinarily undertaken by academic social scientists. Indeed, even the largest social survey carried out by the government — the Current Population survey — is too small for the detailed analysis of topics like American Indian fertility or the professionalization of nursing. The public use samples are the only general source af microdata with sufficient cases to study such small population subgroups.

The large scale of the public use samples also makes them the most suitable source of microdata for policy analysis at the state and local levels. Policy analysts have traditionally focussed on short-run change, but there is increasing recognition of the need to distinguish long-term secular trends from temporary fluctuations. The public use samples also allow policy analysts to set their investigations of State and local conditions in a comparative national context.

In summary, the decennial enumerations of the population include a great deal of information on demography and socioeconomic structure that can only be taken advantage of through the public use samples. We presently understand just the broad outlines of the social transformation that has taken place since the late nineteenth century; published sources provide only limited information on topics such as fertility behavior, urbanization, immigration, household composition, and occupational structure. The public use microdata series allows the construction of comparable cross-tabulations on a wide range of topics that were not covered by census publications or were incompletely tabulated. Perhaps even more important is the potential for pooled multivariate analyses opened up by the availability of microdata. used in combination, the nine data sets spanning a century of cataclysmic social and economic change will comprise our most important resource for the study of changing social structure.

Integration of the Public use Microdata Series

Despite the enormous potential for time-series analysis of the public use samples, to date only a small proportion of the research based on these data has fully exploited the potential for the study of change over time. Many investigators are using the samples as isolated cross-sections. A preliminary bibliography of recent research using the public use samples compiled by the Social History Data Archives at the University of Minnesota reveals that 178 of 220 studies use only one of the eight public use samples currently available.

It is difficult to use more than one of the public use samples at a time because each sample has a different format, different coding schemes, and different documentation. Six separate research teams have been involved in the creation of the samples, and each of them has had their own ideas on how to organize the data. We are faced with eight different occupational classifications with a total of 3200 different categories, and seven incompatible classifications for variables such as birthplace, household relationship, and institution type. In fact, the only variable that is readily comparable across census years is age, and even there the samples differ widely in treatment of missing, illegible, and inconsistent data and in the coding strategy for the very old. Documentation for the eight existing samples is contained in eight separate volumes totaling about 3000 pages. These volumes are for the most part organized differently from one another, and their treatment of comparability issues is often cursory.

Only for the 1950 and 1970 public use samples — where the record layout and coding schemes were made to be reasonably compatible — has there been substantial multi-sample research. Indeed, most of the research using more than one public use sample has focussed on these two census years. This suggests that the incompatibilities of the other samples have been a significant barrier to research on long term social change.

The incompatibility of the public use samples in their present form means that multi-sample studies require a large initial investment to prepare the data for use. The number of investigators using multiple public use samples is growing rapidly. Most have proceeded by creating a set of special-purpose semi-compatible extracts containing a limited number of variables and minimal documentation. This ad hoc approach has already led to increasing duplication of effort. Moreover, given the complexity of the files and the often subtle differences among them, the potential for error is large.

The Social History Research Laboratory plans to convert the public use samples for 1880, 1900, 1910, 1940, 1950, 1960, 1970, 1980, and 1990 into a single consistent format and to prepare an integrated set of documentation oriented to the use of the samples as a series. In the long run, we anticipate adding data for all the remaining census years for which individual-level census enumerations survive; these years are 1850, 1860, 1870, 1920 and 1930. We are currently applying for funding to create a sample for 1920, and plan future applications for the 1850, 1860, 1870 and 1930 census years.

We already have had extensive experience with the entire series of public use samples. Indeed, the creation of common-format extracts of the samples has been a major preoccupation of the Social History Research Laboratory

Table 1

Summary of Availability of Selected Variables: Public Use Samples, 1880-1990

Blank = variable not available
Y = variable available
C = can be constructed
S = sample-line individuals, 1940 and 1950
5 = five-percent sample only, 1970 pus
15 = fifteen-percent sample only, 1970 PUS
N = Neighborhood samples, 1970 PUS
ST = State samples, 1970 Pus
SM = SMSA samples, 1970 PUS
a = "A" sample, 1980 PUMS
b = "B" sample, 1980 PUMS
c = "C" sample, 1980 PUMS

	1880	1900	1910	1940	1950	1960	1970	1980	1990
Geographic Information									
State	Y	Y	Y	Y	Y.	Y	N,ST	Y	Y(1)
urban/Rural residence	Y	Y	Y			Y	N,ST	c	Y
Farm identifier(2)	C	Y	Y	Y	Y	Y	Y	Y	Y
Large cities	Y	Y	Y	Y	Y		SM	a,b	Y
Modified SMA	C	C	C						
SMA				Y	Y				
SMSA							SM	a,b	Y
County or county group	Y	Y	Y	Y	Y		SM	a,b	Y
Personal Characteristics									
Age	Y	Y	Y	Y	Y	Y	Y	Y	Y
Sex	Y	Y	Y	Y	Y	Y	Y	Y	Y
Race	Y	Y	Y	Y	Y	Y	Y	Y	Y
Marital Status(3)	Y	Y	Y	Y	Y	Y	Y	Y	Y
Household Relationahip	Y	Y	Y	Y	Y	Y	Y	Y	Y
Duration of current Marriage		Y	Y		S(4)				
Age at First Marriage				S		Y	5	Y	
Number of Marrbages			Y	S(5)	S(5)	Y	5	Y	
Married in past year?	Y	Y	Y	C,S	C,S	C	C	C	
children ever born		Y	Y	S	S	Y	Y	Y	Y
children surviving		Y	Y						
Surname code	Y		Y	Y	Y				
Subfamily relationships	Y	C	C	Y	Y	Y	Y	Y	Y
secondary fam. relationships	Y	C	C	Y	Y	Y			
Ethnicity and Migration									
Birthplace (country, state)	Y	Y	Y	Y	Y	Y	Y	Y	Y
citizenship/Naturalization		Y	Y	Y	Y		5	Y	Y
Parental birthplace (country)	Y	Y	Y	S	S	Y	15		
Parental birthplace (state)	Y	Y	Y	S	S				
Residence five years ago				Y		Y	15	Y	Y
Year of immigration		Y	Y				5	Y	Y
Mother tongue			Y	S	S	Y	15	Y	Y
speaks English?		Y	Y					Y	Y
Spanish surname	Y			Y	Y	Y	Y	Y	Y

Table 1 (continued)

	1880	1900	1910	1940	1950	1960	1970	1980	1990
Economic status and Employment									
wage and aalary income				Y	Y	Y	Y	Y	Y
Total income					Y	Y	Y	Y	Y
occupation	Y	Y	Y	Y	Y	Y	Y	Y	Y
Industry	C	C	Y	Y	Y	Y	Y	Y	Y
Home ownership		Y	Y	Y		Y	Y	Y	Y
Mortgaged?		Y	Y					Y	Y
Rent/Home value				Y		Y	Y	Y	Y
class of worker			Y	Y	Y	Y	Y	Y	Y
Period worked in census year				Y	S	Y	Y	Y	Y
Hours worked last week				Y	Y	Y	Y	Y	Y
Period unemployed	Y	Y	Y	Y	S			Y	
Year last worked						Y	Y	Y	Y
Currently unemployed			Y	Y	Y	Y	Y	Y	Y
Education and veteran status									
school enrollment	Y	Y	Y	Y	S	Y	Y	Y	Y
Can read	Y	Y	Y						
can write	Y	Y	Y						
Years of schooling				Y	S	Y	Y	Y	Y
Veteran Status			Y(5)	S	S	Y	15	Y	Y

1. Not all geographic information indicated will be available for all versions of the 1990 sample.

2. Definition of farm varies.

3. The "separated" category of marital atatus is not available before 1950; however, the similar category of married, spouse absent can be constructed for all census years.

4. Duration of current marital status.

5. The 1940 and 1950 censuses indicated whether married more than once.

6. Civil war veterans only.

over the past five years. These files are custom designed to meet the research and teaching needs of Minnesota faculty and graduate students. Increasingly, we have been receiving requests for common-format extracts from investigators at other institutions. We currently prepare about 25 common-format extracts a month for a broad range of users.

In the course of our work, we have become intimately familiar with the intricacies of the public use samples. Our staff has invested hundreds of hours in the reconciliation of variables such as occupation and birthplace. It has become obvious, however, that our current procedures — which are duplicated at various institutions across the country — are highly inefficient. What is needed is a complete reworking of all the existing public use samples into an integrated format with complete documentation. This would allow most users to construct their own specialized extracts, and thus dramatically reduce the costs of research.

We are presently in the process of developing a detailed prospectus for the design of such an integrated public use microdata series. It is our hope that prospective users of the data series will provide us with as much feed-back as possible before the design is cast in stone. Copies of the prospectus are available upon request.

REFERENCES

Abel, Marjorie and Nancy Folbre (1990). "A Methodology for Revising Estimates: Female Market Participation in the U.S. Before 1940." *Historical Methods* 23: 167-178.

American Economic Association (1899). *The Federal Census*. New York: Macmillan.

Anderson, Margo J. (1988). *The American Census: A Social History*. New Haven: Yale University Press.

Anderson, Michael (1971). *Family Structure in Nineteenth Century Lancashire*. Cambridge, England: Cambridge University Press.

Barron, Hal (1984). *Those Who Stayed Behind: Rural Society in Nineteenth-Century New England*. New York: Cambridge University Press.

Bose, C. (1987). "Devaluing Women's work: The Undercount of Women's Work in 1900." In C. Bose, R. Feldberg, and N. Sokoloff (eds). *Hidden Aspects of Women's Work*. New York: Praeger, 95-115.

Coale, Ansley J. and Melvin Zelnik (1963). *New Estimates of Fertilbty and Population in the United States*. Princeton, NJ: Princeton University Press.

Conk, Margo A. (1980). *The United States Census and Labor Force Change*. Ann Arbor: UMI Research Press.

_____ (1981). "Accuracy, efficiency, and bias: The interpretation of women's work in the U.S. Census of Occupation, 1890-1940." *Historical Methods* 14:55-72.

Edwards, A.M. (1943). "Comparative occupation statistics for the United States, 1870 to 1940." *Sixteenth Census of the United States: 1940, Population*. Washington, D.C.: U.S. Government Printing office.

Folbre, Nancy and Marjorie Abel (1989). "Women's Work and Women's Households: Gender Bias in the U.S. Census." Social Research 56: 545-569.

Forster, colin and G.S.L Tucker (1972). *Economic Opportunity and White American Fertility Ratios, 1800-1860*. New Haven: Yale University Press.

Goldin, Claudia (1980). "The Work and Wages of Single Women, 1870-1920." *Journal of Economic History* 40: 81-88.

_____ (1983). "The Changing Economic Role of Women: A Quantitative Approach." *Journal of Interdisciplinary History* 12: 707-733.x

Graham, Stephen N. (1980). 1900 Public Use sample: User's Handbook. Seattle: Center for Demography and Ecology, University of Washington.

Hogan, Dennis P., 1981. Transitions and Social Change: The Early Lives of American Men. New York: Academic Press.

Jaffe, A.J. (1956). "Trends in the participation of women in the working force. " Monthly Labor Review 79(5): 559-565.

Katz, Michael B., Michael J. Doucet, and Mark Stern (1982). The Social Organ ization of Early Industrial Capitalism. Cambridge, Mass.: Harvard University Press.

Mcclelland, Peter D. and Richard J. zeckhauser (1982). Demographic Dimensions of the New Republic: American Interregional Migration, Vital Statistics, and Manumissions, 1800-1860. New York: Cambridge University Press.

Modell, John, Frank Furstenberg and Theodore Hershberg (1975). "Social Change and the Transition to

Adulthood in Historical Perspective," Journal of Family History 1: 7-32.

Openheimer, Valerie K. (1970). The Female Labor Force in the United States: Demographic and Economic Factors Governing its Growth and Changing Composition. Population Monograph 5, University of California, Berkley.

Rindfuss, Ronald R. and James A. Sweet (1977). Postwar Fertility Trends and Differentials in the United States. New York: Academic Press.

Ruggles, Steven (1988). "The demography of the unrelated individual, 1900-1950." Demography. 25:4.

_____ (1991). "Comparability of the Public Use Files of the U.S. Census of Population. Social Science History.

Ruggles, Steven and Mbriam King (forthcoming) Fragmentation of the Family: Living Arrangements in America, 1580-1980.

Ruggles, Steven and Russell R. Menard (1990). "A public use sample of the 1880 Census of Population." Historical Methods 23: 104-115.

Stevens, David (1991). "Life-course transitions to adulthood, 1900-1970." Journal of Family History.

Strong, Michael A., samuel H. Preston, Ann R. Miller, Mark Hereward, Harold R. Lentzner, Jeffrey R. Seaman, Henry C. williams (1989). User's Guide: Public Use sample, l9lO Census of Population. Philadelphia: Population Studies Center, University of Pennsylvania.

U.S. Bureau of the Census (1954). The 1960 Census of Population and Housing: 1960. Two National samples of the Population of the United States. Washington, D.C.: U.S. Government Printing Office.

_____ (1972). Public Use samples of Basic Records From the 1970 Census Description and Technical Documentation. Washington D.C.: U.S. Government Printing Office.

_____ (1973). Technical Documentatbon for the 1960 Public Use sample. Washington, D.C.: U.S. Government Printing Office.

_____ (1982). Public Use samples of Basic Records From the 19BO Census: Description and Technical Documentation. washington, D.C.: U.S. Government Printing Office.

_____(1984a). Census of Population, 1940: Public Use Sample Technical Documentation. Washington, D.C.: U.S. Government Printing office.

_____(1984b). Census of Population, 1950: Public Use Sample Technical Documentation. Washington, D.C.: U.S. Government Printing Office.

_____(1989). 1990 Census of Population and Housing. Tabulation and Publication Program. Washington, D.C.: U.S. Government Printing office.

Yasuba, Yasukichi (1963). Birth Rates for the white Population of the United States. Baltimore: Johns Hopkins University

[1] Presented at the IASSIST 91 Conference held in Edmonton, Alberta, Canada. May 14 - 17, 1991.

S. Ruggles, Department of History, University of Minnesota, Minneapolis, MN 55455
ruggles@legohead.hist.umn.edu
(512) 624-4081

Integrating Data Resources and Library Resources: The SPIRES Experience

by Slavko Manojlovich
Memorial University of Newfoundland
St. John's, Newfoundland

INTRODUCTION

The CANSIM database, maintained by Statistics Canada, contains over 400,000 socio-economic time series. The CANSIM University Base, a subset of the main CANSIM base, contains approximately 35,000 numeric time series and is made available to academic institutions on magnetic tape for instructional and research purposes. Prior to 1991, access to the CANSIM University Base at Memorial was provided by the Economics Department via a locally developed FORTRAN program which retrieved time series from tape for specified databank numbers. In 1991 the Library, for budgetary and other reasons, took over the CANSIM subscription. CANSIM has since been loaded into SPIRES, the library's database management system, which is also used to maintain the library catalogue and other library and departmental databases. CANSIM is now accessible to the university community using the same interface as other library databases. It has also been linked to the library catalogue for the display of holdings information associated with related print publications. The remainder of the article will describe various issues regarding loading a numeric database in a library environment.

SELECTING A DATABASE MANAGEMENT SYSTEM

The SPIRES database management system, developed by Stanford University, has been used for over 15 years to manage a variety of information resources including bibliographic, numeric, full-text and image databases. Memorial has been using SPIRES for several years to maintain its library catalogue. Various components of an integrated library system including acquisitions and cataloguing were developed locally, whereas, circulation was obtained from Rensselaer Polytechnic Institute, a member of the SPIRES Consortium. In addition to the library catalogue Memorial has mounted a number of locally developed and commercial databases using the FOLIO interface available in SPIRES (Screen No. 1).

SPIRES has a powerful set of development tools which enable you to tailor the system to accommodate a variety of data types. Since CANSIM was the first numeric database which the library was loading into SPIRES an important first step in the loading process was to identify the DBMS functionality which was required in order to support adequate access to a numeric database. The following functions, although not critical for bibliographic database support, were important for accessing CANSIM:

1) Since end-users typically retrieve time series data for specified time periods SPIRES must prompt for the time series start and end dates.

2) SPIRES must format the data for tabular display or for input to statistical analysis software.

3) SPIRES must output data to a file on the mainframe or on the end-user's microcomputer.

THE CANSIM DATABASE IN SPIRES

The CANSIM file supplied to Memorial contains 1.5 million card images. The logical record describing one time series is comprised of various fixed fields (codes and text strings) and a variable number of data values (Figure 1). SPIRES provides a utility for loading data in its original form thereby alleviating the need to write loader programs. No problems were encountered in loading the CANSIM file. SPIRES has a library of processing functions which allows you to read in data in any form, store it as you like and output it in any form. This inherent flexibility of SPIRES made it easy to implement prompting for start and end dates and the various output formats. SPIRES was also able to accommodate the requirement to output a variable number of data values for specific years when retrieving weekly time series. The various features of CANSIM in SPIRES are illustrated in the following sample search session (please refer to screen displays at the end of the article):

1) A variety of **HELP** screens navigate users through a search. (Screen No. 2).

2) Users can search CANSIM directly using the FIND command on various indexes (Screen No. 3). Terms can be combined from various indexes using boolean operators. The **BRIEF** record display output from a keyword search on **"NEWFOUNDLAND AND WOMEN AND UNEMPLOY#"** is illustrated in Screen No. 4.

3) The **FULL** record display (Screen No. 5) includes in

the **SOURCE** field not only source publication information as supplied by CANSIM but also information regarding Memorial's holdings for that title. For each displayed record SPIRES looks up the holdings information in the library catalogue and includes it in the source field. The addition of holdings information enables the user to easily consult the corresponding print publication for additional information describing the time series or to obtain older data not included in CANSIM. SPIRES performs a corresponding look-up for users searching the library catalogue and notifies them of related information in the CANSIM file (see the **NOTES** field in the record display from the library catalogue — Screen No. 6).

4) Users have the option of displaying data in tabular form (Screen No. 7) or as raw data formatted for input to a variety of time series analysis softwarepackages (Screen No. 8). The user selects the desired format by issuing a **DIS TABLE** or **DIS DATA** command. SPIRES will then prompt the user for the time series start and end dates. Output in either tabular or raw data format may be directed to a file on the mainframe or to the user's microcomputer by issuing either the **SAVE TABLE** or **SAVE DATA** commands. In **SAVE** mode SPIRES prompts the user for the output file name.

5) Users who are uncertain of the appropriate search terms may use the SPIRES **BROWSE** command to scan entries in various indexes (Screen No. 9 and No. 10). Users can selectively display time series from the hit list of displayed terms. All text indexes are may be browsed in the CANSIM application.

6) The above session used SPIRES menu-driven Folio **GUIDED** mode. A **COMMAND** mode option is also available for more experienced users.

THE INTEGRATING ROLE OF SPIRES

The concept of integration in library automation literature typically refers to linkages between various modules in an integrated library system. The CANSIM application described above illustrates how SPIRES expands upon the traditional meaning of integration. Integration in a SPIRES environment also includes:

1) Interface Integration: the use of a common user interface for accessing avariety of databases.

2) Database Integration: the provision of automatic linkages between databases thereby expanding the user's knowledge base (eg. source publication link between CANSIM and the library catalogue).

3) System Integration: the output of data in a variety of formats for direct input to other systems (eg. SAS).

4) Workstation Integration: support for saving data to a file on the mainframe or on the user's workstation.

All of the above further integrate the user into his/her information environment.

CD-ROM VERSION OF CANSIM

Statistics Canada is also distributing a larger portion of the CANSIM database in CD-ROM format. Although CD-ROM is an excellent cost-effective medium for the distribution of large quantities of data it falls short in terms of "integration"as described above, especially in a university environment. The CD-ROM medium forces the user to learn a new user interface. It is not directly linkable to the library catalogue and other campus databases. Access to the database may be restricted to a single workstation. If the CD-ROM is mounted on a campus network, access may be restricted to users with particular hardware. Remote access to a CD-ROM by an end-user working on an old VT-100 terminal may not be possible at all. Migrating data from the CD-ROM to the end-user's statistical analysis software (which is probably on a mini or mainframe) may also be quite cumbersome.

CONCLUSION

Thanks to the integrating power of SPIRES, access to CANSIM at Memorial University is the same as accessing the library catalogue or any other bibliographic database. Based on the success of mounting CANSIM, Memorial is planning on loading census and other fact databases in SPIRES thereby further expanding access to the world of numeric data resources for the library user.

NOTES

1. Additional information on SPIRES is available from: SPIRES Consortium Office, Stanford University, Stanford, California.

[1] Presented at the IASSIST 91 Conference held in Edmonton, Alberta, Canada. May 14 - 17, 1991.

Slavko Manojlovich, Assistant to the University Librarian for Systems and Planning, Memorial University of Newfoundland, St. John's, Newfoundland.

Figure 1

SAMPLE CANSIM RECORD AS SUPPLIED BY STATISTICS CANADA

```
ADD     B      1        90-11-02        195319901210 5 2 310 0 1    199999999*
(12X,  4F17.0)
BANK OF CANADA ASSETS AND LIABILITIES, WEEKLY SERIES (101-103)AND MONTHLY SERIES
 (1-3), WEDNESDAYS AND AVERAGE OF WEDNESDAYS, UNADJUSTED, MILLIONS OF DOLLARS.
B.OF C-STATEMENT/AVE  TOTAL ASSETS  DOLLARS     SCALAR FACTOR 06
SOURCE               BANK OF CANADA REVIEW
AVERAGE OF WEDNESDAYS  CANSIM SERIES IDENTIFIER  000911.1
NOTE                 DATA PUBLISHED IN THE BANK OF CANADA REVIEW APPROXIMATELY 30
CALENDAR DAYS AFTER END OF REFERENCE MONTH.
B      1   1          2348.          2318.          2332.          2352.
B      1   2          2354.          2352.          2410.          2408.
B      1   3          2371.          2364.          2429.          2444.
B      1   4          2390.          2404.          2355.          2357.
B      1   5          2427.          2431.          2309.          2284.
B      1   6          2326.          2304.          2382.          2420.
B      1   7          2369.          2226.          2278.          2310.
B      1   8          2316.          2357.          2433.          2468.
B      1   9          2464.          2476.          2532.          2547
B      1  10          2509.          2368.          2421.          2472.
B      1  11          2467.          2511.          2528.          2531.
B      1  12          2519.          2543.          2550.          2571.
B      1  13          2514.          2406.          2429.          2492.
B      1  14          2519.          2580.          2604.          2629.
B      1  15          2632.          2645.          2696.          2670.
B      1  16          2606.          2540.          2574.          2646.
B      1  17          2652.          2719.          2800.          2855.
B      1  18          2885.          2997.          2956.          2951.
B      1  19          2800.          2753.          2768.          2809.
B      1  20          2838.          2857.          2857.          2928.
B      1  21          2880.          2848.          2943.          2869.
B      1  22          2822.          2728.          2736.          2816.
B      1  23          2830.          2842.          2902.          2905.
B      1  24          2860.          2895.          2950.          2927.
B      1  25          2906.          2824.          2876.          2896.
B      1  26          2920.          2909.          2981.          2998.
B      1  27          3030.          3066.          3064.          3066.
B      1  28          3062.          2940.          2990.          3075.
B      1  29          3105.          3227.          3242.          3309.
B      1  30          3178.          3205.          3215.          3221.
B      1  31          3136.          3012.          3072.          3167
                          .....
                          .....
                          .....
```

Screen No. 1
MENU OF PUBLIC ACCESS DATABASES AT MEMORIAL

Folio contains 17 files. Press the RETURN key to see the rest of the list.

Public Information Files

1.	BIBLIO	MUN Library Online Catalogue
2.	CANSIM	Canadian Socio-Economic Time Series Database
3.	CNS ARCHIVES	Archival records of the Centre for Nfld. Studies
4.	EXTENSION	Division of Extension Resource Library
5.	LABBIB	Canadian Labour Bibliography
6.	LINSIC	Labrador Institute of Northern Studies Info Cen.
7.	MICROLOG	Canadian Research and Report Literature
8.	MUNFLA	MUN Folklore and Language Archive
9.	OEIC	Ocean Engineering Information Centre
10.	PAB	Newfoundland Periodical Article Bibliography
11.	SLIDES	Grenfell College Fine Arts Slides Database

===> Press the RETURN key to see the rest of the list,
or select a file by typing its name or number
WHICH FILE? CANSIM

Screen No. 2
INITIAL CANSIM HELP SCREEN

-CANSIM selected

CANSIM is Statistics Canada's computerized data bank and information and retrieval service. The CANSIM University Base on FOLIO contains 32,425 of the most popular time series in the main base. Subject areas covered by CANSIM include system of national accounts, population, labour, prices and international trade. The database is currently updated once a year (the last update was December, 1990). Time series may be output as either tabular displays or in raw data format suitable for input to a variety of time series analysis software packages. Raw data or tables may be output to a file at NLCS or downloaded to the user's microcomputer using KERMIT communications software. The CANSIM FULL display includes the MUN library holdings of the associated print publications. Enter HELP CANSIM OUTPUT for more information on CANSIM display formats. For additional information contact Joy Tillotson, Information Services, QE II library, ext. 7427.

For more information on this file:	type HELP.
To search this file:	type FIND.
To search headings in sequence:	type BROWSE.
To select a different file:	type SELECT.
To see all your options:	type OPTIONS.

YOUR RESPONSE: FIND

Screen No. 3
MENU OF CANSIM INDEXES

You can search the CANSIM file for any of the following information:

Type of Search		Example
MATRIX	matrix numbe	911 or 000911
SOURCE	keyword index to source publications	BANK OF CANADA REVIEW
MAWORD	keyword index to matrix titles	BANK OF CANADA ASSETS
SEWORD	keyword index to series titles	mortgage#
KEYWORD	keyword index to titles and notes	consumer price
FREQ	frequency	MONTHLY
DB	databank numbers	B10

Indicate below the type of search you want by typing the name or names of the type of information you have, e.g. MATRIX.
Use the BROWSE command to examine entries in various indexes.

TYPE OF SEARCH: FIND WORD NEWFOUNDLAND WOMEN UNEMPLOY#

Screen No. 4
CANSIM BRIEF RECORD DISPLAY

CANSIM / Search: Find KEYWORD NEWFOUNDLAND WOMEN UNEMPLOY#
Result: 6 series
1)
Series: UNEMPLOYMENT RATE WOMEN 20-24 YRS. (monthly, 1975-1990) (Databank No: D774089) Matrix: NEWFOUNDLAND, BASIC LABOUR FORCE CHARACTERISTICS, MONTHLY FROM JAN 75, UNADJUSTED (RAW) IN THOUSANDS. SELECTED SERIES ARE LINKED TO PREVIOUS SURVEYS OF JAN 66 OR JAN 70. (No: 002078)
2) Series: UNEMPLOYMENT RATE, 25 YRS AND OVER, WOMEN (monthly, 1975-1990) (Databank No: D772686) Matrix: NEWFOUNDLAND, BASIC LABOUR FORCE CHARACTERISTICS, MONTHLY FROM JAN 75, UNADJUSTED (RAW) IN THOUSANDS. SELECTED SERIES ARE LINKED TO PREVIOUS SURVEYS OF JAN 66 OR JAN 70. (No: 002078)

Series continue; press RETURN to see next page
To see a full series: type DISPLAY FULL followed by a number.
To begin a new search: type FIND or BROWSE.
To select a different file: type SELECT.
To get more information: type HELP or OPTIONS.
YOUR RESPONSE: DF 1

Screen No. 5
CANSIM FULL RECORD DISPLAY

```
CANSIM / Search: Find KEYWORD NEWFOUNDLAND WOMEN UNEMPLOY#
Result: 6 series
Series 1
Series: UNEMPLOYMENT RATE WOMEN 20-24 YRS. (monthly, 1975-1990)  (Databank No:
        D774089)
Matrix: NEWFOUNDLAND, BASIC LABOUR FORCE CHARACTERISTICS, MONTHLY FROM JAN 75,
        UNADJUSTED (RAW) IN THOUSANDS. SELECTED SERIES ARE LINKED TO PREVIOUS
        SURVEYS OF JAN 66 OR JAN 70. (Matrix No: 002078)
Source: MONTHLY LABOUR FORCE DATA (71-001), STC
        MUN HOLDINGS: The labour force. La population active 71-001, LOCATION:
        MUGD, HOLDINGS: v. [10-11]-[33]- 1954-
Notes: SOURCE MATERIALS COVERING BACKGROUNDS OF PREVIOUS LFS REVISIONS AND
        MODIFICATIONS OF DEFINITIONS, OF CONCEPTS AND OF LFS DESIGN, AS THEY
ARE          INTRODUCED IN THE CURRENT REVISION MAY BE OBTAINED FROM THE
DIVISION.          REQUESTS SHOULD REFER TO 1) 'LABOUR FORCE INFORMATION', CAT.
NO. 71-001P,            FEB76 2) 'RESEARCH PAPER' #2 AND #3 3) 'METHODOLOGY OF
THE  CANADIAN LABOUR             FORCE', STATISTICS CANADA, CAT. NO. 71-526,
OTTAWA, 1976
        Scalar Factor: 00   Data Output Format: (10X, 4F17.1)
        Missing Values = 9999999999. or equiv.    Secure data = all asterisks.
```

Screen No. 6
LIBRARY CATALOGUE DATABASE RECORD DISPLAY

```
BIBLIO / Search: Find TITLE LABOUR FORCE
Result: 16 titles
Title 16
          TITLE: The labour force. La population active
   FORMER TITLE: Labour force bulletin. Main d'oeuvre bulletin 1945-19
      PUBLISHED: Ottawa, Statistics Canada, Labour Force Survey Division
    DESCRIPTION: v
          DATES: v.[1]- 1945-
           ISSN: 03806804
          NOTES: V.32, no. 1-2 not published
                 Vol. numbering begins with v.6, no.1, Mar. 1960
                 Issues for 1945-49 called no.1-13
          NOTES: Vols. for 1945-1971 issued by the Dominion Bureau of
                 Statistics. ; 197 -19 by Statistics Canada, Labour Force
                 Surveys Section
     SUBJECT(S): Labor supply—Canada—Statistics—Periodicals.
  OTHER ENTRIES: Statistics Canada. Labour Force Surveys Section
                 Canada. Dominion Bureau of Statistics
 OTHER ENTRIES:  Labour force bulletin
   CALL NUMBER:  71-001, LOCATION: MUGD, HOLDINGS: v. [10-11]-[33]-   1954-
           RSN:  75348973
==>       NOTES: *** NOTE:  Numeric data associated with this record are
```

Screen No. 7
CANSIM TABULAR RECORD DISPLAY

1)
Series: UNEMPLOYMENT RATE WOMEN 20-24 YRS. (monthly, 1975-1990) (Databank No: D774089)
Matrix: NEWFOUNDLAND, BASIC LABOUR FORCE CHARACTERISTICS, MONTHLY FROM JAN 75, UNADJUSTED (RAW) IN THOUSANDS. SELECTED SERIES ARE LINKED TO PREVIOUS SURVEYS OF JAN 66 OR JAN 70. (No: 002078)

Year	Jan	Feb	Mar	Apr
1985	32.7	30.0	29.2	31.0
Year	May	Jun	Jul	Aug
1985	32.7	33.8	25.8	27.0
Year	Sep	Oct	Nov	Dec
1985	25.4	26.0	25.9	23.3

.....
.....
.....

Screen No. 8
CANSIM RAW DATA OUTPUT FORMAT

Series: UNEMPLOYMENT RATE WOMEN 20-24 YRS.
Matrix: NEWFOUNDLAND, BASIC LABOUR FORCE CHARACTERISTICS, MONTHLY FROM JAN 75, UNADJUSTED (RAW) IN THOUSANDS. SELECTED SERIES ARE LINKED TO PREVIOUS SURVEYS OF JAN 66 OR JAN 70. (Matrix No: 002078)

(10X, 4F17.1) (monthly, 1985-1990, Scalar Factor: 00)

32.7	30.0	29.2	31.0
32.7	33.8	25.8	27.0
25.4	26.0	25.9	23.3
26.9	25.1	28.0	26.1
25.9	25.9	30.1	23.8

.....
.....
.....

Screen No. 9
MENU OF CANSIM BROWSE INDEXES

```
You can browse the following types of information in the CANSIM file:

 Type of Browse                                Example

   MATI  matrix titles                         Bank of Canada
   SETI  series titles                         Short term loans
   SOPH  source publication                    BANK OF CANADA REVIEW
   WORD  list of all terms in text fields      consumer

 Indicate below what you wish to browse by typing the name of the type of
 information you have, e.g. SOPH.

 Use the FIND command to perform a direct search.

 TYPE OF BROWSE: BROWSE WORD FISH
```

Screen No. 10
CANSIM BROWSE KEYWORD INDEX

```
CANSIM / Search: Browse WORD FISH
Result filed under the following headings:

-3) Word: FIRMS  (690 series)
-2) Word: FIRST  (4328 series)
-1) Word: FISCAL  (129 series)
0) Word: FISH  (332 series)
1) Word: FISHERY  (1 series)
2) Word: FISHING  (3688 series)
3) Word: FITTING  (2 series)
4) Word: FITTINGS  (7 series)
5) Word: FIVE  (926 series)
6) Word: FIXED  (1053 series)
7) Word: FIXTURE  (63 series)
8) Word: FIXTURES  (103 series)

_____Headings continue; press RETURN to see next page_____
To see a heading's series:    type DISPLAY followed by a heading number.
To see a full series:         type DISPLAY FULL followed by a number.
To begin a new search:        type FIND or BROWSE.
To get more information:      type HELP or OPTIONS.
YOUR RESPONSE: DIS 1
```

LISA

Library & Information Science Abstracts

- **international scope and unrivalled coverage**
  LISA provides English-language abstracts of material in over thirty languages. Its serial coverage is unrivalled, 350 titles from 60 countries are regularly included and new titles are frequently added
- **rapidly expanding service which keeps pace with developments**
  LISA is now available monthly to provide a faster-breaking service which keeps the user informed of the rapid changes in this field
- **extensive range of non-serial works**
  including British Library Research and Development Department reports, and conference proceedings
- **wide subject span**
  from special collections and union catalogues to word processing and videotex, publishing and reprography
- **full name and subject indexes provided in each issue**
  abstracts are chain-indexed to facilitate highly specific subject searches
- **available in magnetic tape, conventional hard-copy format, online (BRS, ORBIT and Dialog file 61) and now on Silver Platter CD-ROM**
  Twelve monthly issues and annual index

Subscription: UK £186.00
Overseas (excluding N. America) £217.00
N. America US$429.00

Write for a free specimen copy to

Sales Department
Library Association Publishing
7 Ridgmount Street
London WC1E 7AE
Tel: 01 636 7543 x 360

LA

CURRENT RESEARCH
in library & information science

- **CURRENT RESEARCH** is an international quarterly journal offering a unique current awareness service on research and development work in library and information science, archives, documentation and the information aspects of other fields
- **The journal provides** information about a wide range of projects, from expert systems to local user surveys. FLA and doctoral theses, post-doctoral and research-staff work are included
- **Each entry provides** a complete overview of the project, the personnel involved, duration, funding, references, a brief description and a contact name. Full name and subject indexes are included
- **Other features include** a list of student theses and dissertations and a list of funding bodies. Each quarter, an area of research is highlighted in a short article

CURRENT RESEARCH is available on magnetic tape, as well as hard copy, and can be searched on DIALOG

Subscription: UK £97.00
Overseas (excluding N. America) £113.00
N. America US$224.00

Write for a free specimen copy to

Sales Department
Library Association Publishing
7 Ridgmount Street
London WC1E 7AE
Tel: 01 636 7543 x 360

LA

Conference Announcement and Call for Papers

IASSIST 1992

DATA, NETWORKS AND COOPERATION: LINKING RESOURCES IN A DISTRIBUTED WORLD

The 18th Annual Conference of the International Association of Social Science Information Service and Technology (IASSIST) will be held at the Concourse Hotel in beautiful Madison, Wisconsin, from Tuesday, May 26, through Friday, May 29, 1992. IASSIST brings together individuals engaged in the acquisition, processing, maintenance, and distribution of computer-readable text and numeric social science data. Founded in 1974, the membership includes data archivists, librarians, information specialists, social scientists, researchers, planners, and government agency administrators from around the world.

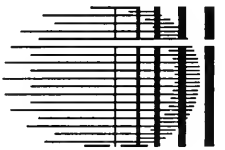
The central conference theme expresses IASSIST members' concern for managing and sharing computer-readable data during a time of increasing demand coupled with decreasing fiscal resources. The theme touches upon the need for institutional cooperation as well as careful planning to meet the needs of the future. The conference program features workshops, contributed papers, roundtable discussions and poster sessions reflecting international viewpoints on these concerns. The Program Committe is soliciting proposals in areas including:

- New cooperative data ventures through network technology
- Management of data library collections
- Archiving of electronic records at federal levels and below
- Process and development of data collections
- Data Library hardware and software issues
- Moving to UNIX
- Copyright and computer files
- Coping with fiscal restraint
- Major comparative data sources
- Hypercard
- Major comparative data sources
- The producers of data
- Utilization of new technologies
- Text Encoding Initiative
- Developing data centers
- Disaster management
- Integrating data services with traditional library services

Proposals for presentations of any kind should be submitted to the Program Committee Chair by December 15, 1991. Proposals should be accompanied by brief abstracts (ca. 200 words).

For further information, contact the Program Committee Chair:

Ilona Einowski
Data Archivist
Data Archive and Technical Assistance
University of California, Berkeley 2538 Channing Way
Berkeley, CA U.S.A. 94720
(415) 642-6571
CENSUS85@UCBCMSA.BITNET



IASSIST

INTERNATIONAL ASSOCIATION FOR SOCIAL SCIENCE INFORMATION SERVICE AND TECHNOLOGY

• • • •

ASSOCIATION INTERNATIONALE POUR LES SERVICES ET TECHNIQUES D'INFORMATION EN SCIENCES SOCIALES

Membership form

The International Association for Social Science Information Services and Technology (**IASSIST**) is an international association of individuals who are engaged in the acquistion, processing, maintenance, and distribution of machine readable text and/or numeric social science data. The membership includes information system specialists, data base librarians or administrators, archivists, researchers, programmers, and managers. Their range of interests encompases hard copy as well as machine readable data.

Paid-up members enjoy voting rights and receive the **IASSIST QUARTERLY**. They also benefit from reduced fees for attendance at regional and international conferences sponsored by **IASSIST**.

Membership fees are:
Regular Membership. $20.00 per calendar year.
Student Membership: $10.00 per calendar year.

Institutional subcriptions to the quarterly are available, but do not confer voting rights or other membership benefits.

Institutional Subcription:
$35.00 per calendar year (includes one volume of the Quarterly)

I would like to become a member of IASSIST. Please see my cholce below:

- ☐ $40 Regular Membership
- ☐ $20 Student Membership
- ☐ $70 Institutional Membership

My primary Interests are:

- ☐ Archive Services/Administration
- ☐ Data Processing
- ☐ Data Management
- ☐ Research Applications
- ☐ Other (specify)_____

Please make checks payable to IASSIST and Mail to :

Ms Kay Worrell
Treasurer, IASSIST
% The Conference Board
845 Third Avenue
New York, NY 10022-6601

Name / title

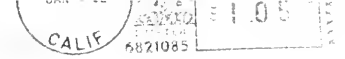
Institutional Affiliation

Mailing Address

City

Country / zip/ postal code / phone

ISSR
UNIVERSITY OF CALIFORNIA
405 HILGARD AVENUE
LOS ANGELES, CA 90024-1484



SERIALS DEPT.
U. OF N. CAROLINA-CHAPEL HILL
CB#3938 DAVIS LIBRARY
CHAPEL HILL, NC 27599

BOOK RATE